

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: Ed Vesna - Italy

Searcher Phone #: 308-4501

Searcher Location: Biotec Lib.

Date Searcher Picked Up: 5/7/01

Date Completed: 5/9/01

Searcher Prep & Review Time: _____

Clerical Prep Time: 5 min

Online Time: 2 min

Type of Search

NA Sequence (#) 5

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbit _____

Dr. Link _____

Lexis/Nexis _____

Sequence Systems AB5503

WWW/Internet _____

Other (specify) _____

This Page Blank (uspto)

STIC-Biotech/ChemLib

41 777

From: Goldberg, Jeanine
Sent: Friday, May 04, 2001 4:55 PM
To: STIC-Biotech/ChemLib
Subject: 09/619,643- maize est

1. please search SEQ ID NO: 1-5

THANKS

Jeanine Enewold Goldberg
1655
CM1--12D11
Mailbox-- 12E12
306-5817

This Page Blank (usp10)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:01:46 ; Search time 2791.02 Seconds
(without alignments)
2266.853 Million cell updates/sec

Title: US-09-619-643-1
Perfect score: 429
Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaagaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 737392952 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_da1:*
17: em_da2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	33.6	1213	13	AF332876 Oryza sat
2	136.8	31.9	1254	15	WHTWZFLB D16416 Wheat mRNA
3	136.8	31.9	2862	15	WHTWZFLB D16415 Wheat gene
4	50	11.7	33818	3	MTCY159 Z83863 Mycobacteri
5	49.6	11.6	1250	85	AB000714 Homo sapi
6	49.6	11.6	1601	85	AF007189 Homo sapi
7	49.2	11.5	7764	95	RN081010 AF007189 Rattus norv
8	48.4	11.3	19712	3	SCD65 AL392176 Streptomy
9	48	11.2	139999	12	AC018727 Oryza sat
10	47.8	11.1	34893	3	SC17 AL096743 Streptomy
11	47.8	11.1	51575	68	AC023810 Mus muscu

REFERENCE 1 (bases 1 to 33818)
AUTHORS Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
Tejeda,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Felwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REMARK Erratum: [[Published erratum appears in Nature 1998 Nov
12;396(6707):1901]]
REFERENCE 2 (bases 1 to 33818)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1781046.
COMMENT Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in t3parse (Kirogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
FEATURES
source 1..33818
Location/Qualifiers
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
misc_feature <1..513
/note="Fragment designated v009. Does not represent a
physical clone"
complement(171..9380)
/gene="fas"
complement(171..9380)
/gene="fas"
/note="RV2524c, (MTCY159_32-MTV009_09c), len: 3069. fas
Fatty Acid Synthase, similar to eg G580746 fatty acid
synthase (3104 aa), fasta scores, opt: 3867, E(): 0
(50.0% identity in 3032 aa overlap), contains PS00017
ATP/GTP-binding site motif A, PS00606 Beta-ketoacyl
synthases active site"
/codon_start=1
/transl_table=11
/product="fas"
/protein_id="CAB06201.1"
/db_xref="GI:3261686"
/db_xref="SPTREMBL:P95029"
/translation="MTIHDRVSADRGDSPPHTHALVDRLMAGEPYAVARGGOGSA
WETLEELYSATGITEFLATLGEAEILLDPYTDLIYRPTGEPFLQWVRLAEDP
VPSDKHLTSANVSVPGLVLTQATRALARQMDLVATPPVAMACHSGCVLAELAKA
GARDELVALAQILGAAGTVLARRGISVLGDRPMVSVTNADPERIGRLDEDFAPV
GTVVPELALINGRRRAVITGTPQLSFEFLCYKQISEKEDARKNKVRGDFVSPV
FEPVQVEGFTPLRLSDGIDIVAGWAERKAGDALARELADAILIRKVDWDEITRVH
AAGARWIDLPGDILTRLTAVIRGIGIVAPADRGQRNLFTVGATPEVARAMS
YAPTVVRLDGRVKLSTFKTRLTGSPILLAGMPTTYDAKIVAAANAGHWAELAGC

GOVTEIFGNRIEQMAGLLEBGRTYQFNALFLDLYLWKLOVGRKLVOKAROSGAID
GVNVSAGIPDDDEAVLELIDELDGISIHVVKPGTIEIRSVIAIEVPRKPIIMVY
EGGRASGHWSMEDDLDLATYIESRANATKVCVGGIGIPRAAEVNEVPRKPIIMVY
FLPMIDIDILVGTATAMATKESTSPSVRYMLVDQGDQWISAKRAGGVMSSSQIG
ADIHEDIDLSRCGRLLDEVAAGDAVAREDDITAMAKAKRKYDGMADMTLOHL
RRYVELAIGENSTADTASVCSPMIADTWRPFEDMLORARLHPDQEGTOTLFTD
AGLNDPOQAIATALLARYPDATVOLHAPADVPFVCTCKTGKRVNPEVVIDVRM
WRSLSMOAHBARLDADAVCIIPGASVAGITTRDEPVGILLDEFEQAEIVDVGAGV
EPKDVASRLGRADVAGPLVAVLADPVRNARVTVNVRHIADEPAEQVNDGCEPNR
ATHSGTARLTHGDGVALSVPDGSVDWIDFETPLPANTVDGFTVPIATEDTSAMRY
LATIAGVDSPEFLPAVANGTATITVDYMDREHVAADHTQATFEGEPLASLTNVDAVY
GPCMPAVPAATIGSAVTTDGERVYEGSLSHLDHARAVGOLPTVPRAQITATNANA
TDTDMGRVPAVSVAVTGADVIAATLEERFILLRTGSAELADPARAGASANAATDT
PRRRRDVTITAPVDMRPFAVSGDNHPIHDRAALLAGLESPVHGMMLLSAAOHA
VTATDQARPAIARLVGMPARFGLGWRPDEVDREVERIGDGAETIDVAVARDAVIF
MSASRLAAPTVAAPFPGCOIOHGMKEVARSARKAARKVDMDTDFPDRLTGSLVH
VVRNDPTSIISGVYHHDPDVLVLTQOTVQAMATVAAAOVAKREBCAPFEGAATAG
HSVGYTALACYTGITQLEALLBNYFHSGSKMDIVPRDELGRSNYRLAIRPQIDIL
DSVDPAVPAVIAESTGEFLITVFNLSQVYALAGVIRGLEALAEVERRELITGGR
RSFLLVPGIDVPEHSRVLRVGVAEFRRLSDRVMRPDADPDLLIGRYIPNLVRLFTLD
RDFIDEIRDLVPAEPLDEILDYDTWLERREMARVFIILLAMOFASPVAMETOD
LLFTEAAGGGLVERFEIYKSSPTVAGLATNTLKPEYASHSTVEVLANRDAVIF
ATDTPDEPEDEPVAESPAPDVYSEAPVAPAASSAGPRDDLVPAADATATLTA
LSAKRIDIOTEIELSISTDQASGRRLQILVDGSELNICAIDGAEAPDAGLRQVY
TKLARTYPRGPVLSDAINDQIRYTLGSGRRPRAIARVYKKTWELBGRKHTVER
ALGTREGSSVRGAGHLHEGALADAVDVDAVANAVARGVSELPSALSGSGG
ATIDAAALSEFTDQITGREGVLASARVILQGLDGVNALPAPPSLIDLYTAEL
GADPRLVAPVDFPKAVFDDRMASAREDIYKMLDDEGIDIDWPLERFEGAGH
VADQATWMOCKSLAAGROHASLYGRARAAKNEPPEPYGEVAVVYVAGSKSIAS
VVARLDGGAATVITATSKLDEERLAFYETLVRDHAARVGAALMIYAAVNAASVDDVY
EMIGREPESIGPOSITHKDKAOTPTLTPPAPRYGDSLTSAGRAEMKVIAMAYO
RLIGLSTIGERDLASRLHYLRPSPRKMGFGDGAIGKSKALDAVYSWAHESM
AARVSLAHLIGWRTGTLMDHNDALIAVAEAGATYITSTEMAALLDLDLDAESKA
AARSPKIALDGLGAEANIDAELEAKAREGMSAAVYDEDEAPGATVLELSPRRG
TPAPRPQMDLDVDPADLVIVGAEITPYGSSVTRFEMEYENELSAIGVLELMTTG
LIRMPDDPOCPWYDYESGEMVDESELYORHYDAVAVORGIPEFDVDAIDGHSPL
VSYLEKDFARFVYSEADARAFAVEEDPTIRVPDSTDMOVIKRGKETIRVVRKTK
LSRVYGGQITPGEFPTWGISADNAGSTDRILAVNMVATVYPAFLSSGSPRAVRYVH
PSLVANTGTGKGGSMTQWTHGMLLRKPNDFQCVLITNIIAHHVOSYVSYGA
MHPVAACATAVSVEEGVDRILRGKQDLVVAAGDILTLGIIIGFVGMATADTSMK
GHRPHADSKFSVNRRLRGVEAGGKDSPLIARLALRMLGVLVAVARAQSGFTG
VHTSIAPAGLALGARGGKDSPLARALAKGAADVAVASKDSTLADNPTEL
HERLADALGRSGEAPLFVYSSKSLTGAKGAAYVOMMGICOLIRDQVYIPNRSDCY
DDELAGSNHPVWVYRDTLRLGKTFPLKAKMLTSLGCHVSGIYALVHPOATFASIDPAO
RADTORRADARLALGQRRLASLIAGGAPMTPRGDRRDRHHAPEQDASMLNPAAR
LGDEGAYYIG"
383..>33818
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y159"
complement(1200..1250)
/gene="fas"
complement(1200..1250)
/note="PS00606 Beta-ketoacyl synthases active site"
complement(3120..3143)
/gene="fas"
complement(9390..9393)
/note="possible RBS, GCAG, for RV2524c"
complement(9900..10622)
/gene="RV2525c"
complement(9900..10622)
/gene="RV2525c"
/note="RV2525c, (MTCY159_31), len: 240 aa. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV2525c"
/protein_id="CAB06183.1"
/db_xref="GI:1781077"
/db_xref="SPTREMBL:P95028"
/translation="MSVSRDVLKPPATPGVLGVAVASSLRAAPASAGSLGTLDY
AAGVPAISORIAAGAVGAIIRVYSDRRPGAMWGKIPOLSSARDLSAGSLGTCYQY
GKGSTADMLGGSAGVQHARRGSELHAAAGGPTSNPIYASIDNPNSTEQYKNQIVPL

Query Match	Best Local Similarity	Score	DB	Length
Matches 86: Conservative	0: Mismatches	60: Indels	0: Gaps	0: Gaps
QY 40	ccaagaagaagcgtctcctgcgcgcgtctcgcattcccaagcagacgtagagcttgcgcgt	11.7%	DB 3:	33818:
Db 33021	ccgaccgccgctggtgcatcaccggccgctgattcattcaacagatgtctgcgtccggga	58.9%	Pred. NO. 11:	
QY 100	gcgcgtacctccatgagactactccacgcgcgtcatcactgagactgcagacagacgacg	0: Mismatches	60: Indels	0: Gaps
Db 33081	ccgggtgtagctgattaccgaagaccgaacatggtctccacaaacttgccgccgcagcagc	0: Mismatches	60: Indels	0: Gaps
QY 160	accaccaccaccaccaccaccagcagc	0: Mismatches	60: Indels	0: Gaps
Db 33141	gccaccaccaccaccaccagcagcagc	0: Mismatches	60: Indels	0: Gaps
RESULT 5	AB000714	1250 bp	mRNA	PRI
LOCUS	AB000714	1250 bp	mRNA	PRI
DEFINITION	Homo sapiens hrvp1 mRNA for RVp1, complete cds.	27-OCT-1997		
ACCESSION	AB000714			
VERSION	AB000714.1	GI:2570128		
KEYWORDS	RVp1.			
SOURCE	Homo sapiens tissue_l1b:lung cdna to mRNA.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
	Primates; Catarrhini; Hominiidae; Homo.			

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1250)	Katahira, J.	Direct Submission	
Submitted (26-JUN-1997)			to the DDBJ/EMBL/Genbank databases. Jun
Katahira, Institute for Microbial Diseases, Osaka University,			Department of Bacterial Toxinology, 3-1, Yamadaoka, Suita, Osaka,
565, Japan (E-mail: katahira@iken.osaka-u.ac.jp, Tel: 81-6-879-8285,			Fax: 81-6-879-8283)
2 (sites)	Katahira, J., Sugiyama, H., Inoue, N., Horiguchi, Y., Matsuda, M. and		
Sugimoto, N.			Clostridium perfringens enterotoxin utilizes two structurally
related membrane proteins as functional receptors in vivo			J. Biol. Chem. 272 (42), 26552-26558 (1997)
97476271			
FEATURES	SOURCE	location/Qualifiers	
	1..1250	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/tissue="lung"		
	199..861		
	/gene="hrvp1"		
	199..861		
	/gene="hrvp1"		
	/product="RVp1"		
	/protein_id="BAA22986.1"		
	/db_xref="GI:2570129"		
	/translation="MSMGIRITGALAVIGMIGTVCCALPMRVSAFIGSNITTSN		
	IMEILMANCVYSGOMCKYVDSLLALPDLDQARALLTYAILILAFGLLVALVQ		
	CTNCVQDDIKAKITIVAGVFLALTLTPVSSANITIIIDFINFVPEQRRKNG		
	AGLVGNMAAALQLLGGALLCCSCPREKRTYATKVVYSAFSTGPGASLGTGRD		
	YV		
	1250		
	/note="13 A nucleotides"		
	BASE COUNT 202 a 452 c 409 g 187 t		
	ORIGIN		
	polya_site		
	Query Match 11.6%, Score 49.6, DB 85; Length 1250;		
	Best Local Similarity 54.3%; Pred. No. 55;		
	Matches 100; Conservative 0; Mismatches 84; Indels 0; Gaps 0;		
OY	153 gagagcgacacacaccaccacacacagagagagcgccaccgctcaccagcgcgagca	212	
DB	868 GACGCGAGGAGACCCACCCACACACACACACACACACACACACACACACCGAGCTG	927	
OY	213 gggcgccacactgtcccccggggtggcgcaagaagaagcgctcggcgcgcaagcgctcag	272	
DB	928 GAGCGCGACACACGAGCGCATATCCAGCGTGCAGCGCTGCGCTGGAGGCCAGGCCACCCCCAGAA	987	
OY	273 aggaaggaacactcgcgctgtgctgtcatatgctctcccgcgcgagccagccagcgctc	332	
DB	988 GCCACGAGACCCCGCGCTGGACTGGGGGCGAGCTTCCACGACGCGCTTGGCGGCC	1047	
OY	333 aggc 336		
DB	1048 GGCG 1051		
RESULT	6		
AF007189	1601 bp	DNA	PRI 19-NOV-1998
LOCUS			
DEFINITION	Homo sapiens claudin 3 (CLDN3) gene, complete cds.		
ACCESSION	AF007189		
VERSION	AF007189.1		
KEYWORDS	GI:2459927		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1601)		
AUTHORS	Peacock, R.E., Keen, T.J. and Inglehearn, C.F.		

TITLE	protein	Analysis of a human gene homologous to rat ventral prostate.1
JOURNAL	Genomics	46 (3), 443-449 (1997)
MEDLINE	98110580	
REFERENCE	2 (bases 1 to 1601)	
AUTHORS	Keen, T.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-JUN-1997)	Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath St., London EC1V 9EL, UK
FEATURES	Location/Qualifiers	
source	1. .1601	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7q11"	
mrna	<477. .>1139	
	/gene="CLDN3"	
	/product="claudin 3"	
gene	<477. .>1139	
	/gene="CLDN3"	
	477. .1139	
CDS	/gene="CLDN3"	
	/note="similar to RVP.1; similar to Rattus norvegicus ORF encoded by GenBank Accession Number M74067"	
	/codon_start=1	
	/product="claudin 3"	
	/protein_id="AAC78272.1"	
	/db_xref="GI:2459928"	
	/translation="MSMGLEETITGALVLCGLTIYCCALPMMRYSAFISNITTSNON	
	IMELWMNVVOSTGMOCKVYDLSLALPDQLNARALIVAILLAFLGLVLAIVGAD	
	CTNCVDDDTAAKTTIVAGVLFLLAALLTVPSWMSANTIIIDYFNBPVPEAKRREGK	
	AGLVGNMAAALQLLGGALLCCSCPPREKKTATKVVYSAPRSTPGASLGTGYDRKD	
	YV"	
polya_signal	1508. .1513	
BASE COUNT	251 a 595 c 520 g 235 t	
ORIGIN		
Query Match	11.6%; Score 49.6; DB 88; Length 1601;	
Best Local Similarity	54.3%; Pred. No. 49;	
Matches 100; Conservative	0; Mismatches 84; Indels 0; Gaps 0;	
OY	153	gagagcagcacccaccaccacacacacagcagcagcgcgcacgcgtcaccagcgtgcaga 212
Db	1146	GACGCAGGAGAGACCCACACACACACACACACACACACACACACACACCGAGCTG 1205
OY	213	gggcgcaccacgtccgcacgggggtgggcacaagaagaagcgtccgcgcgcacgcgtccag 272
Db	1206	GAGCGCGCAGCCAGCGGCATCCAGCGCGTGCCTGGAGAGCCAGCCACACCCACAGA 1265
OY	273	aggagaggaacactcgctgcgtctgcgtctcattcctccgcgcgcgcgcgcgcgcgc 332
Db	1266	GCCACGAGAGCCGCCCGCTGGACTGGGGCAGCTTCCACGACGACGACGCTTTGGCGGCC 1325
OY	333	aggc 336
Db	1326	GGGC 1329
RESULT	7	
LOCUS	RNU81010	7764 bp DNA ROD 31-MAY-1997
DEFINITION	Rattus norvegicus glutamate receptor kainate (GRK5) gene, promoter region, 5' untranslated region, partial sequence and intron 1, complete sequence.	
ACCESSION	U81010	
VERSION	U81010.1	GI:2138291
KEYWORDS		
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	

[illegible]

REFERENCE 3 (bases 1 to 19712)
AUTHORS Cerdano,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire
CB10 1SA E-mail: barrellesanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are
numbered using the following system eg SC7B7.01c, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an upstream
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid D65 lies
between and overlaps cosmids D6 and D69 on the AseI-D genomic
restriction fragment.
Location/Qualifiers
1. 19712
/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"
/complement(1..440)
/gene="SCD65.01c"
1. 19712
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid D65"
1. 120
/note="nominal overlap with Streptomyces coelicolor cosmid
SCD6"
complement(<1..440)
/gene="SCD65.01c"
/note="SCD65.01c, possible lipoprotein (fragment), len:
>146 aa; similar to TR:Q55943 (EMBL:D64005) Synecocystis
sp. hypothetical 22.9 kDa protein SLI0788, 196 aa; fasta
scores: opt: 203 z-score: 217.8 E(): 0.00012: 33.9%
identity in 118 aa overlap. Contains correctly situated
match to Prosite entry PS00013 Prokaryotic membrane
lipoprotein lipid attachment site and possible N-terminal
region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative lipoprotein (fragment)"
/protein_id="CAC08365.1"
/db_xref="GI:10178341"
/translation="MRTAVVAAGAALVLAAGSGGDSAGHGHAATAPSSAPS
SASASASASASOGHAAADYAFKGMIPHRQAKVENADLAPGRARSSEVQKLAADIKK
AOPPIRTLSGMLTSMGEDIYPAEGAMDSHTSDMSGGMWTAEE"

misc_feature complement(276..338)
/gene="SCD65.01c"
/note="alanine and serine amino acids rich region"
complement(381..413)
/gene="SCD65.01c"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(583..987)
/gene="SCD65.02c"
/gene="SCD65.02c" (987)
complement(583..987)
/gene="SCD65.02c"
/note="SCD65.02c, possible secreted protein, len: 134 aa.
Contains possible N-terminal region signal peptide
sequence. High alanine and glycine amino acid residues
content"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC08366.1"
/db_xref="GI:10178342"
/translation="MTRSTGSSRPAGRGFLALLVLAAGVAGMGLACAMPQAG
AGHGYWTADGVPHADGGCTHTDGSSSHLDADGTCAAGTGSAYTPPALGALLD
APAAPAPGAFPPAPHDGRAPDLAEIQLRTI"
1010..1480
/gene="SCD65.03"
1010..1480
/gene="SCD65.03"
/note="SCD65.03, unknown, len: 156 aa. Contains match to
Prosite entry PS00038 Myc-type, 'helix-loop-helix'
dimerization domain signature"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC08367.1"
/db_xref="GI:10178343"
/translation="MVRPKRPSPGHANTHOEVPVSHPRLPAPPEALADAKRLSLPRI
VVTCGSRREMTEDVRETTGRIVVKKPCDKSPHALMSDPVPAEALKARLDELH
RAKIRLADVLVYGVGDVYGDSTRAEIYARSIDKPVFTIRPVDGNAVERGRP"
1133..1180
/gene="SCD65.03"
/note="PS00038 Myc-type, 'helix-loop-helix' dimerization
domain signature"
complement(1498..2055)
/gene="SCD65.04c"
complement(1498..2055)
/gene="SCD65.04c"
/note="SCD65.04c, possible tetr-family transcriptional
regulatory protein, len: 185 aa; similar to SW:BE1.RHME
(EMBL:U39940) Rhizobium meliloti regulatory protein Bel1,
203 aa; fasta scores: opt: 193 z-score: 251.5 E():
1.5e-06; 27.3% identity in 183 aa overlap. Contains Pfam
match to entry PF00440 tetr, Bacterial regulatory
proteins, tetr family and possible helix-turn-helix motif
at residues 33..54 (+3.34 SD)"
/codon_start=1
/transl_table=11
/product="putative tetr-family transcriptional regulatory
protein"
/protein_id="CAC08368.1"
/db_xref="GI:10178344"
/translation="MTGSTRGPPMDPLRRERILDAALDVVAEHGAIKVTFRRIAAGV
PLGSIVYPPDMQHLTLDTATRALEVSTRYGLLLAQTROGAEEAVYIYCGKRWG
TDRNLILSTELVAFATRHPEVRAVRSWMAKSSLSARHPDLTARALDLIEGFSIH
NSVLSLPTAREDDVAAYSAVVDHRR"
complement(1867..2007)
/gene="SCD65.04c"
/note="Pfam match to entry PF00440 tetr, Bacterial
regulatory proteins, tetr family, score 44.00, E-value
9.7e-10"
complement(2056..2059)
RBS 2170..2173
RBS 2181..3428
gene /gene="SCD65.05"

CDS		2181..3428 /gene="SCD65_05" /note="SCD65_05, possible integral membrane protein, len: 415 aa; similar to TR:P75810 (EMBL:AB000186) Escherichia coli MscK protein, 402 aa; fasta scores: opt: 1317 z-score: 1364.7 E(): 0; 54.1% identity in 392 aa overlap. Contains possible hydrophobic transmembrane spanning regions"
gene		/codon_start=1 /transl_table=1 /product="putative integral membrane protein" /protein_id="CAC08369.1" /db_xref="GI:10178345" /translation="MTDRARVORRRRLATLEFFFLNGIAMSSWVTPTDIRDLGVSTAQ MGTLFELSGSMGTGILCSRLVRFGTTPVTAAGTLIIYASVAVWGSAALASAPLV TAGLCEFGVGAGEVALINVDGAVERITGPVP.LTHGGCGSLGVLAGAMATAAPA AFAPHMLISAVVAAYACIIFTFYAIGAVAGCIRAPRAPSAGSKSVQVADRLLILI GATLYAMAALAEAGANDPLPILMNDGHGLDAAGSLYYVGRPAANTLRFGTFPELSRR SRATVVASASVSGALGLCLVIFSDNTVFAAAAVLFKLGSGLFPVALSADGSPAEK TARGLVATIGIYLAFLVGPLGLGDHYGLRPAMVVYLAIFAANAAPAAPAGRGRSR HHGADTRTPAEPAPERSAORTREG"
CDS		3425..3574 /gene="SCD65_06" /length=129 /note="SCD65_06, hypothetical protein, len: 49 aa; highly similar to the C-terminal region of Sw:SUG_ECOLI (EMBL:X69949) Escherichia coli SUG protein, chaparrone, 105 aa; fasta scores: opt: 108 z-score: 175.6 E(): 0.028; 45.7% identity in 46 aa overlap. Contains match to Prosite entry PS00012 Phosphopantetheine attachment site"
Query Match	11.3%, Score 48.4; DB 3; Length 19712;	
Best Local Similarity	48.8%; Pred. NO. 26;	
Matches 159; Conservative	0; Mismatches 166; Indels 1; Gaps 1;	
Oy	87	gagctgcgcytgctgcggaactcattgatggatccacgcgcgtcacatgatgagctgc 146
Db	5929	GACCCTGCCGCCGCCTCCGATCTCATCTGTCGCCCGCAGAACATCGGGCGGCGAGACTGG 5870
Oy	147	gaagaagaagacgac 206
Db	5869	GATGAACGGCAATGTTCCACCAAGATCAACCCGAGACCACTGGCCCCCATGCTCTCACCGG 5810
Oy	207	cgaagaggcgccccaccctgcccccaggggtggccaagaagaagcgctcgcgcgccagcg 266
Db	5809	CATCGAGGGCGTTCCTCCCTGACACAAGGTGCCCCCGCTGGAACGTGTCGTCGTGACC -CG 5751
Oy	267	tctcaagagaaagaaacactcgcgctgtgacctgatactctcccgcgcgagacgaccaac 326
Db	5750	ACTACGGTGCTCAAACTCCGATCCGATGACCGTGGTGCACCCCCTGCGCGGCATCGCGAGG 5691
Oy	327	gcgtccagcg 386
Db	5690	AGGCGGTGGCGCGCGCGCGCGCGCGCTTCATATGTGTCCACGCGTCCCGAACCGCGCGCGG 5631
Oy	387	aatccttcgctctctaccagcgctg 412
Db	5630	ACGCGTTCACAGCTATACAGCGTGTG 5605
RESULT 9		
LOCUS	AC018727/c	DNA
DEFINITION	Oryza sativa chromosome 10 BAC OSJNBa0056g17 genomic sequence,	PLN 05-JAN-2001
ACCESSION	AC018727	
VERSION	AC018727.10	GI:12039362
KEYWORDS	HTG.	
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa	
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza	

REFERENCE (bases 1 to 139999)
AUTHORS Buell,C.R., Yuan,Q., Moffett,K.S., Hill,J.N., Burr,P.C., Hsiao,J., Zisman,V., Pal,G., Bowman,C.L., Fujii,C.Y., Vankken,S.E., Bowman,C.L., Craven,B., Utteback,T.R., Khalak,H., Feldblum,T.V., Quackenbush,J., White,O., Salzberg,S.L. and Fraser,C.M.
TITLE Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence
REFERENCE
AUTHORS 2 (bases 1 to 139999)
TITLE Buell,R.
JOURNAL Direct Submission
SUBMITTED (17-DEC-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 139999)
AUTHORS Buell,R.
JOURNAL Submitted (05-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT On Jan 5, 2001 this sequence version replaced gi:12025621.
Address all correspondence to:rice@tigr.org

BAC clone OSJNBa0056G17 is from Oryza sativa chromosome 10
The orientation of the sequence is from spe to t7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GenScan and GenScan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/genemark/), Pgenes
(http://www.softberry.com/), and GeneSplicer (Mihaila Pertea and
Steven Salzberg, contact mperle@tigr.org), searches of the
complete sequence against a peptide database and the plant EST
database at TIGR (http://www.tigr.org/cdb/eg1.shml). Annotated
genes are named to indicate the level of evidence for their
annotation. Genes with similarity to other proteins are named after
the database hits. Genes without significant peptide similarity but
with EST similarity are named as unknown proteins. Genes without
protein or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Atlan Smil,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1. 139999
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="10"
/map="C405"
/clone="OSJNBa0056G17"
1. >213
/gene="OSJNBa0056G17.16"
1. 213
/gene="OSJNBa0056G17.16"
/note="similar to cellulose synthase 8 GB:AAF89968
GI:9622888 (Zea mays)"
1. 213
/gene="OSJNBa0056G17.16"
/codon_start=1
/product="putative cellulose synthase 5-partial"
/protein_id="AAG4616.1"
/db_xref="GI:12039380"
/translation="KILGVFEFWVAHLHYPAFKGLMGRGRPTIVYVWSGLVAIT
ISLIWAIKPPSADNSOLGGSFSP"
complement(915..1010)
/rpt_family="(GGG)n"
complement(join(1122..1267,1540..1704,1788..1867,
2028..2131,2443..2544,3960..4113,4204..>5940))
/gene="OSJNBa0056G17.1"
complement(1122..5940)
/gene="OSJNBa0056G17.1"
/note="EST BE039949 from this gene"
complement(join(4045..4113,4204..5940))

/gene="OSJNBa0056617.1"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAG46167.1"
 /db_xref="GI:12039381"
 /translation="MLHRTPTPRRLSAAAAAFAETTPRPSHPPEPPPRLSPLRYDA
 TVSRSLALALITFEFLMCAFCAPFAEHPSSSDRLPAARLASLRTFAPIILHOLRA
 GCPTRPHFLLLRILRYMGQIYPLVQLFQDMPLMGHPNAPANNVYLDVLRTRHH
 SALLCRNPSPNLTITAILILTHLCRSGNMWVACPLRGLHGFILPSALNVAEFC
 CTKLAASTELLGLFTLVSGYQLTSAAMCTLIARLCREGLEDIAIMAKMLASGSP
 PVVAVYITVVALRYAGRHDIATETELFASMSSTNSPDLVLYNIMDMCTDKRDAL
 GYVLLHJSHSGKMPDAVYLTSTLVRALHLSRNSYLPRFLIDSAIDIPDYLAQNSVNL
 RKSQSPSQAQNFITNMTKCDIRPDSYSVYGLDLCQLGRDHAINYHSIVSSDPS
 NAYVAAILCGVKGKGNRMALMLNEAVRQNTALDVACTYVHLGHLQAHLLIEAQM
 LEFKRRSGMASNTCTYINMLRGLCTRDEHAWLWFLREMECSDEVENDSISYNIYLVE
 LIKLIQHISSATLALANIENHDSSTNDLLIACAS"
 5604..5785
 repeat__region
 /rpt_family="(CGGGGG)n"
 5835..5921
 /rpt_family="(CGG)n"
 complement(<8345..>9535)
 /gene="OSJNBa0056617.1"
 complement(8345..9535)
 /gene="OSJNBa0056617.1"
 /note="predicted by genemark.hmm"
 complement(8345..9535)
 /gene="OSJNBa0056617.1"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG46168.1"
 /db_xref="GI:12039382"
 /translation="MADSHLDGVPAMRKTTVSSSSSLNIAFLSMVATNLLSTYHST
 RASTAPLLPLPAHQADQDELRLQTLAIATVQSQNLHRSSTPPRPPPELLYS
 RLAPLASCSAHPDLHFMSTYEPSPCPDPLSLAPRLHLCGHPILRRRCSSSRI
 SSSHNLNLPDPSPLPDLVAVRMPGAKCPSCPLPSIGDLAPREARFLARBP
 LDITYPOLLRLASLSRAQPIRLGLDVGCGTTLAARKLRANATVLTITNMLGPSV
 AAAGVPLPLRAIDQRFPGDATMDVLRAQHANRMIPKAEFLIYWDADRYLPRGQ
 LQK"
 repeat__region
 9206..9290
 /rpt_family="(CGGGGG)n"
 complement(join(<8691..10099,10826..10955,11282..11459,
 11659..11768,11903..12018,12376..12637,12766..13075,
 13214..>13329))
 /gene="OSJNBa0056617.2"
 complement(9691..13329)
 /gene="OSJNBa0056617.2"
 /note="EST D47052, AU0966093, AU075883 from this gene"
 complement(join(9913..10099,10826..10955,11282..11459,
 11659..11768,11903..12018,12376..12637,12766..13075,
 13214..13297))
 /gene="OSJNBa0056617.2"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AAG46171.1"
 /db_xref="GI:12039385"
 /translation="MAVDILNHLIANGVGEAATACGLRTLYRNNAOLVYSIGILY
 AADKIVDEAFPAASIKRPSALFAMEGCVALLFLPRLSANGFAPDPATVFIHRLP
 LFEVSLVTLAVADVSPASAKILFTREGGASLIVAGYTLSTRIRYKQULIPA
 BPMKRSPFGLEFRMAANAVASFAVAIYSPFNLGTTATTCDFELPSNVFGITLGS
 RLPSQVKKVLPHTICCALADLANAVAGYLSRSGVDVNLGDTYFESPLASVAGDITLGS
 FLGYSIISFAFSMFRNORLIVRRHAAEIFTSLVASTFSLGALIGVLEPLEPILIS
 ILPRITVIALRLVSLDEGVNTSVAAYVLTGLINAFNAQVMDKLRLEKDPARI
 GTSASAHGLGAAVSAKREPALPFCALATVNLGTVAASLPCSLAIRISLIVFIAGDASA
 SQTQHS"
 repeat__region
 complement(10278..10350)
 /rpt_family="AT-rich"
 complement(12207..12227)
 /rpt_family="AT-rich"
 complement(join(<14187..14282,14405..14515,14635..14703
 14841..14996,15085..15127,15235..15330,15805..15918,
 16055..>16368))
 /gene="OSJNBa0056617.15"

	gene	complement(14187..16368) /gene="OSUNBa0056G17.15" /note="highly similar to cholinephosphate cytidylyltransferase GB:CAA70317 GI:1657382 (Pisum sativum); EST C26593 from this gene" complement(join(14187..14282,14405..14515,14635..14703, 14881..14966,15085..15127,15235..15330,15805..15918, 16059..16243)) /gene="OSUNBa0056G17.15" /codon_start=1 /product="putative cholinephosphate cytidylyltransferase" /protein_id="AAG646173.1" /db_xref="GI:12039387"
CDS	repeat_region	/translation="MARVSSSRKTRAPSNSKSPISQTPPELITDRFVYADGI FDLPHGARLEDAQAKLLFPNTYLVLGCCDELTNRKGTAVTODERYSLRHKWY DEVLPDAPWLVTQFEIDKHIDIVAAHDLADTSAGANOVFEVKIKGFKEPTKD GVSDILMRILKDYNOYVMRNLRGRTRDKLDGYVEKOLOVNMKIIGLFRTVAH OEKIQTYAKTAGIINHEEMLANADRNVAGPIELEKECHNNMETAIKGRIQELKRQISR GITGLMQDPYAA"
gene	repeat_region	complement(16596..16625) /rpt_family="CARGCG.n" /rpt_family="TAA)n" complement(16673..16698)
mRNA	repeat_region	/rpt_family="CARGCG.n" complement(join(17509..17640,18143..18245,18370..18437 18565..18722,19375..19564,19831..>20874)) /gene="OSUNBa0056G17.30" complement(17509..20874) /gene="OSUNBa0056G17.30"
CDS	gene	/note="similar to 4 coumarate CoA ligase GB:AAC24503 GI:3258635 (Populus tremuloides)" complement(join(17509..17640,18143..18245,18370..18437, 18565..18722,19375..19564,19831..20874)) /gene="OSUNBa0056G17.30" /codon_start=1 /product="putative 4-coumarate CoA ligase" /protein_id="AAG646175.1" /db_xref="GI:12039389"
		/translation="MOPDAATATHPSSLFSYSAAATGLYSLHPPLPLPSDSPSLVPHL FSLPHHHSHLVNDAPPAATACADCFRLVSSLAAGLRRLHTARGSVLPLPNSLT PPVFALVIATGAATAMPNSSAPAEIALTAIRDPAPSLVASTHNAAKBLPLNLSLT VPPOFOOHDDPDDEFPHALLFTPDRETPMEVNGOQDOAAVLYISSGGSGSKVY VTHNHLTAMVELIPREFASOTTRPRANDNYLALPMHYIGLSLFAVGLSLDCSTVV MRRNVDAVKAIRKYVTLPLVPLPMSALLRNPDELSDILOVSSGAAPLNHLLI
Query Match		11.2%: Score 48; DB 12; Length 139999;
Best Local Similarity	54.5%:	Pred. No. 13;
Matches	96; Conservative	0; Mismatches 80; Indels 0; Gaps 0;
Oy	45	agaagtcgctccgcgcgcgtcgcattccaccagcaacgtatagactgccgtgtcgccgt 104
Db	64903	AGCCCTGCCCATGGCCACCACGACCAAGCCCCACGCCCTCCAGTCGCACACTGGCT 64844
Oy	105	gaattccatggagtacctccaccagccgtcatcatgtgagactgtcgatgcagaagaagaccac 164
Db	64843	CAGACCACCCGAGGCCACCACTCTCTGCCACCGCGATCACCAACGATGCTGAGACAG 64784
Oy	165	caccacaaccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 220
Db	64783	CACCACAGCAGCAAGAACACCACTCCACGACGACACGACACGCCACCGCAGCCACC 64728
RESULT	10	
SC17	LOCUS	SC17 34893 bp DNA BCT 01-JUL-1999
ACCESSION	DEFINITION	Streptomyces coelicolor cosmid 17.
KEYWORDS	VERSION	AL096743.1 GI:5419991
SOURCE		alanine-rich protein; dihydroipicolinate synthase; dihydroxy-acid dehydratase; integral membrane transport protein; large secreted protein; lyase; oxidoreductase; periplasmic substrate binding protein; sugar; transcriptional regulator; transferase; transport associated protein; xylanase; zinc-binding dehydrogenase. Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 1 to 34893)
 AUTHORS Redenbach,M., Kleiser,H.M., Denaplatte,D., Elchner,A., Cullum,J.,
 Kinash,H. and Hopwood,D.A.
 TITLE A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
 MEDLINE 97000351
 REFERENCE 2 (bases 1 to 34893)
 AUTHORS Seeger,K. and Harris,D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 34893)
 AUTHORS Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUL-1999) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK
 COMMENT Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC and Beowulf Genomics
 Details of S coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)
 CDS are numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons
 using a specially developed Hidden Markov Model (Krogh et al.,
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
 program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov.jp/jun/cgt-bin/frameplot.pl>. CAUTION: We may not have predicted the
 correct initiation codon. Where possible we choose an upstream
 codon (atg, gtg, ttg or (att)) which is preceded by an upstream
 ribosome binding site sequence (optimally 5-13bp before the
 initiation codon). If this cannot be identified we choose the most
 upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions.
 Cosmid 17 Cosmid 517 lies between and overlaps with cosmids 139
 and C22 on the AseI-I genomic restriction fragment.
 Location/Qualifiers
 1..34893
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 17"
 complement(1..285)
 /gene="SC17.01c"
 complement(1..285)
 /note="SC17.01c", partial CDS, possible xylanase, len:
 >95aa; similar to the N-terminal region of many eg.
 TR:030426 (EMBL:AF005383) xylanase from the thermophile
 Caldocellum saccharolyticum (1347 aa) fasta scores: opt:
 263, z-score: 357.5, E(): 1.4e-12, (35.2% identity in 91
 aa overlap)."
 /codon_start=1
 /transl_table=11
 /label="SC17.01c"

gene /product="putative xylanase"
 /protein_id="CAB46384.1"
 /db_xref="GI:5419992"
 /translation="MTAVESAGTYYTNVPLNADWSDPDVYVGVGGFYLTASSFGKVP
 LPILASRLVMTLVGNLELEPEREFAARRRHDGCVAPSLRHAGRFMT"
 complement(282..1127)
 /gene="SC17.02c"
 complement(282..1127)
 /note="SC17.02c"
 /gene="SC17.02c", unknown, len: 281aa
 /codon_start=1
 /transl_table=11
 /label="SC17.02c"
 /product="hypothetical protein"
 /protein_id="CAB46385.1"
 /db_xref="GI:5419993"
 /translation="MTGNDSPVLNAGRPVARYVTRPELPRLSPRPYLHPVATLAC
 AVTEPSPADHGHGAVGVADPDVGHNMWGRRTVVRGRPELDNHCORTAQLD
 PDGYVEELRWYAAAGALELRERTYAAAEELGTALADPTFSLTVTRDPLTGS
 GRGAAGGFFWRARKESTPPGVTAADREGREHGRADMLALGTETWSLVFAGAT
 RTRRDPWFVRETEYPGVSSLAHDDRVTAVAGDTLVIRIVVADGRILDGAALVR
 KAVSP"
 complement(1124..2281)
 /gene="SC17.03c"
 complement(1124..2281)
 /note="SC17.03c", possible oxidoreductase, len: 385aa;
 similar to many proteins of undefined function eg.
 TR:045376 (EMBL:X00711) from a locus required for
 lipopolysaccharide biosynthesis in Bordetella pertussis
 (350 aa) fasta scores: opt: 232, z-score: 266.7, E():
 1.6e-07, (28.0% identity in 275 aa overlap). Contains Pfam
 match to entry PF01408 GRO_IDH_MoCA, Oxidoreductase
 family.
 /codon_start=1
 /transl_table=11
 /label="SC17.03c"
 /product="putative oxidoreductase"
 /protein_id="CAB46386.1"
 /db_xref="GI:5419994"
 /translation="MSRPVVLACARGHGNVNRIRLERAGLVRLAGICELPLTE
 QEAGEELPEQADFGALDSTGARVAVTCPTPIPTDLATPARGVHLLKEKPPAS
 VAEFRPMADGAGAGVACGVGFOGSGAIVARIELVARGAIGVVGIGAGAVP
 DYFRAPRSGSKRLNGADYIDGALTNPLAHVATRALAGTARADEYTGITETSHAN
 DIEADDSVNRIGTARGHRTVAATLCKERAEPYVLVHGSSGVETWYKDRDLAR
 GGHGPEEHNDPTLDELVDHLLTGAALVPBETGAFMVVAIRAPADPELPA
 AMRRLPGERRRVPGIDLVAAADTLSTVSELGAPAPRPREHREHVSRT"
 complement(1511..2272)
 /gene="SC17.03c"
 /note="Pfam match to entry PF01408 GRO_IDH_MoCA,
 Oxidoreductase family, score 71.60, E-value 1.7e-17."
 complement(2278..3174)
 /gene="SC17.04c"
 complement(2278..3174)
 /gene="SC17.04c"
 /note="SC17.04c", possible integral membrane transport
 protein, len: 298aa; similar to many eg. SW:Y0RK_BACSU
 hypothetical ABC transporter from Bacillus subtilis (300
 aa) fasta scores: opt: 582, z-score: 650.2, E(): 6.9e-29,
 (31.7% identity in 293 aa overlap) and TR:087858
 (EMBL:AL031013) putative transmembrane sugar transport
 protein from Streptomyces coelicolor (279 aa) fasta
 scores: opt: 540, z-score: 604.3, E(): 2.5e-26, (34.7%
 identity in 277 aa overlap). Contains Pfam match to entry
 PF00528 BPD_transp, Binding-protein-dependent transport
 systems inner membrane component and Prosite match to
 PS00402 Binding-protein-dependent transport systems inner
 membrane comp sign. Also contains possible membrane
 spanning hydrophobic regions."
 /codon_start=1
 /transl_table=11
 /label="SC17.04c"
 /product="putative integral membrane transport protein"

[illegible]

KEYWORDS	REFERENCE	
SOURCE	AUTHORS	
ORGANISM		
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

HTG: HTGS-PHASE1.
house mouse.
Mus musculus
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 51575)
Meltzer,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlack,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonin,D., Brooks,A., Brown,J.,
Bahay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansy,J., Gill,R.,
Corrill,J.H.C., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Kosar,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Hosack,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Martin,R.,
Massey,E., Mcleod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 51575)
Worley,K.C.
Direct Submission
Submitted (18-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
on Nov 3, 2000 This sequence version replaced g1:8248603.

```

Center Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

-----
Project Information
Center project name: MABO
Center clone name: RP23-252B10

-----
Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primed Bodipy; 76% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 42708 bases at least Q40
Consensus quality: 59319 bases at least Q30
Consensus quality: 66755 bases at least Q20
Estimated insert size: 39143; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; sum-of-contigs estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_diff.data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2683: contig of 2682 bp in length
*
* 2683: gap of unknown length
*
* 2783: contig of 3111 bp in length
*
* 5893: gap of unknown length
*
* 5894: contig of 1446 bp in length
*
* 5994: gap of unknown length
*
* 7439: gap of unknown length
*
* 7440: contig of 1767 bp in length
*
* 7540: gap of unknown length
*
* 9306: gap of 1856 bp in length
*
* 9407: gap of unknown length
*
* 11362: gap of unknown length
*
* 11363: contig of 1374 bp in length
*
* 11363: gap of unknown length
*
* 12737: contig of 1116 bp in length
*
* 12836: gap of unknown length
*
* 13952: contig of 1116 bp in length

```


Web site: <http://www.seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L3205
Center clone name: 346_P-7
----- Summary Statistics -----
Sequencing vector: M13: M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181484 bases at least Q40
Consensus quality: 190159 bases at least Q30
Consensus quality: 193745 bases at least Q20
Insert size: 207000; agarose-fp
Insert size: 197427; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1091: contig of 1091 bp in length
* 1092 1191: gap of 100 bp
* 1192 3065: contig of 1874 bp in length
* 3066 3165: gap of 100 bp
* 3166 4305: contig of 1140 bp in length
* 4306 4405: gap of 100 bp
* 4406 5983: contig of 1578 bp in length
* 5984 6083: gap of 100 bp
* 6084 8322: contig of 2239 bp in length
* 8323 8422: gap of 100 bp
* 8423 10098: contig of 1676 bp in length
* 10099 10198: gap of 100 bp
* 10199 12592: contig of 2394 bp in length
* 12593 12692: gap of 100 bp
* 12693 14684: contig of 1992 bp in length
* 14685 14784: gap of 100 bp
* 14785 17163: contig of 2379 bp in length
* 17164 17263: gap of 100 bp
* 17264 19900: contig of 2637 bp in length
* 19901 20000: gap of 100 bp
* 20001 23054: contig of 3054 bp in length
* 23055 23154: gap of 100 bp
* 23155 26011: contig of 2857 bp in length
* 26012 26111: gap of 100 bp
* 26112 29412: contig of 3301 bp in length
* 29413 29512: gap of 100 bp
* 29513 32991: contig of 3479 bp in length
* 32992 33091: gap of 100 bp
* 33092 36913: contig of 3822 bp in length
* 36914 37013: gap of 100 bp
* 37014 40880: contig of 3867 bp in length
* 40881 40980: gap of 100 bp
* 40981 45015: contig of 4035 bp in length
* 45016 45115: gap of 100 bp
* 45116 50321: contig of 5206 bp in length
* 50322 50421: gap of 100 bp
* 50422 54284: contig of 3863 bp in length
* 54285 54384: gap of 100 bp
* 54385 59430: contig of 5046 bp in length
* 59431 59530: gap of 100 bp
* 59531 64373: contig of 4843 bp in length
* 64374 64473: gap of 100 bp
* 64474 70503: contig of 6030 bp in length
* 70504 70603: gap of 100 bp
* 70604 77829: contig of 7226 bp in length
* 77830 77929: gap of 100 bp
* 77930 85976: contig of 8047 bp in length
* 85977 86076: gap of 100 bp

FEATURES

Source
* 86077 91694: contig of 5618 bp in length
* 91695 91794: gap of 100 bp
* 91795 99208: contig of 7414 bp in length
* 99209 99308: gap of 100 bp
* 99309 107499: contig of 8191 bp in length
* 107500 107599: gap of 100 bp
* 107600 115731: contig of 8132 bp in length
* 115732 115831: gap of 100 bp
* 115832 123297: contig of 7466 bp in length
* 123298 123397: gap of 100 bp
* 123398 133979: contig of 10582 bp in length
* 133980 134079: gap of 100 bp
* 134080 144631: contig of 10552 bp in length
* 144632 144731: gap of 100 bp
* 144732 159118: contig of 14387 bp in length
* 159119 159218: gap of 100 bp
* 159219 172437: contig of 13219 bp in length
* 172438 172537: gap of 100 bp
* 172538 200727: contig of 28190 bp in length.

Location/Qualifiers
1. .200727
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/map="11"
/clone="RP23-346P7"
/clone_1lb="RPCT-23 Female Mouse BAC"
1. .1091
/note="assembly_fragment"
1192. .3065
/note="assembly_fragment"
3166. .4305
/note="assembly_fragment"
4406. .5983
/note="assembly_fragment"
6084. .8322
/note="assembly_fragment"
8423. .10098
/note="assembly_fragment"
10199. .12592
/note="assembly_fragment"
12693. .14684
/note="assembly_fragment"
14785. .17163
/note="assembly_fragment"
17264. .19900
/note="assembly_fragment"
20001. .23054
/note="assembly_fragment"
23155. .26011
/note="assembly_fragment"
clone_end:5p6
vector_side:right"
26112. .29412
/note="assembly_fragment"
29513. .32991
/note="assembly_fragment"
33092. .36913
/note="assembly_fragment"
37014. .40880
/note="assembly_fragment"
40981. .45015
/note="assembly_fragment"
45116. .50321
/note="assembly_fragment"
50422. .54284
/note="assembly_fragment"
54385. .59430
/note="assembly_fragment"
59531. .64373
/note="assembly_fragment"
64474. .70503
/note="assembly_fragment"

KW	basic helix-loop-helix zipper, homeotic; homeodomain; homeobox; MADS;
KV	homeodomain zipper; LIM domain; AP2; ERBBS; zinc finger domain;
KW	type 2 Cys2His2; CCAAT box element; MYB; ss.
XX	
OS	Eucalyptus grandis.
XX	
PN	MO200053724-A2.
XX	
PD	14-SEP-2000.
XX	
PF	09-MAR-2000; 2000WO-US06112.
XX	
PR	11-MAR-1999; 99US-0266513.
PR	18-AUG-1999; 99US-0149485.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PA	(FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX	
PI	Wood M, McGrath A, Shenk MA, Glenn M;
XX	
WP	WPI: 2000-579369/54.
XX	
PT	New isolated polynucleotide encoding a plant transcription factor for
PT	producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT	having modified gene expression or modified activity of a polypeptide
XX	
PS	Claim 1; Page 484; 747pp; English.
XX	
CC	The present invention relates to novel plant transcription factors from
CC	Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC	sequence for one such transcription factor. The transcription factor may
CC	be used to produce a plant having modified gene expression such as a
CC	woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC	maioyary species or to modify the activity of a polypeptide in a plant.
CC	The transcription factors of the present invention are members from the
CC	binding families of regulatory proteins: bZIP, bZIP family of G-box
CC	binding factors, basic helix-loop-helix zipper,
CC	homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC	and ERBBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC	and MYB.
XX	
SQ	Sequence 273 BP; 37 A; 124 C; 73 G; 39 T; 0 other:
Query Match	15.9%; Score 68.2; DB 21; Length 273;
Best Local Similarity	60.8%; Pred. No. 2.2e-05;
Matches 161; Conservative	0; Mismatches 78; Indels 26; Gaps 2
OY	131 ccacgcgtacccagcgagcgagggcgccaccatctccccaagggtgggccaagaagc 250
Db	1 cctctctctctgctgctgagctagacactccctcgcgcgagcccttggccaagcgagc 60
OY	251 gctgcgagcgagcagc-----ggtcgaagaggaagaactcgctgtctgtcta 302
Db	61 gtttcaagagccccaccacaacccgcctctcgagagagtaactgcctctgtgata 120
OY	303 tgcctcccgcgagcagccacgcgcttccaaggcgcgcgcgcggtcccgctccgc-- 360
Db	121 tgcctgcgcgagcgagcgagcgagcactaccctccgcgagcctccctccggtctct 180
OY	361 -----ggaagtcaagtgctccgcttcgtgcgggaatccttcgctctacc 404
Db	181 ccgagagcggtcaatgttgctactacagtgcccgcaatcgacagaaggcttccctctacc 240
OY	405 aggcgtggtgggggcaaaagaccag 429
Db	241 aggccttggggcgccacaagcgccag 265

XX	C56651;
AC	
XX	
DT	25-JAN-2001 (first entry)
XX	
DE	Eucalyptus grandis transcription factor DNA sequence #522.
XX	
KM	Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW	poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KM	basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW	homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
KW	type 2 Cys2His2; CCAAT box element; MYB; ss.
XX	
OS	Eucalyptus grandis.
XX	
PN	WO200053724-A2.
XX	
PD	14-SEP-2000.
XX	
PF	09-MAR-2000; 2000WO-US06112.
XX	
PR	11-MAR-1999; 99US-0266513-
XX	18-AUG-1999; 99US-0149485.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	(FLET-) FLETCHER CHALLENGE FORESTS LTD.
P1	Wood M, McGrath A, Shenk MA, Glenn M;
XX	
DR	WPI: 2000-579369/54.
XX	
PT	New isolated polynucleotide encoding a plant transcription factor for
XX	producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT	having modified gene expression or modified activity of a polypeptide
XX	-
PS	Claim 1: Page 484; 747pp; English.
XX	
CC	The present invention relates to novel plant transcription factors from
CC	Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC	sequence for one such transcription factor. The transcription factor may
CC	be used to produce a plant having modified gene expression such as a
CC	woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC	mahogany species or to modify the activity of a polypeptide in a plant.
CC	The transcription factors of the present invention are members from the
CC	following families of regulatory proteins: bZIP, bZIP family of G-box
CC	binding factors, basic helix-loop-helix zipper,
CC	homeo/c/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC	and ERBSs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC	and MYB.
SQ	
	Sequence 310 BP; 42 A; 149 C; 85 G; 34 T; 0 other:
	Query Match 15.3%; Score 65.8; DB 21; Length 310;
	Best Local Similarity 62.9%; Pred. No. 6.4e-05;
	Matches 149; Conservative 0; Mismatches 62; Indels 26; Gaps 2.
QY	219 ccactgcgcccgagggttgcccaagaagacggcgcgcagc-----gtctc 270
DB	3 ccccttcgcgcgaccttggccaagcgaaagctctcaaacgcccccacaaccgcgcc 62
QY	271 aaagagaggaacactcgcgtctgtgcctctaccagcgctggggggccaaagaccag 429
DB	63 cgaggaagcagatcatctgcctcttgtcataatgtctgcgcgcggcgccgcgcgcgac 122
QY	331 ccaagcgccgcgcgcgcgcgtccgcgcgc-----ggagtcaagttg 372
DB	123 cctcccccgcgcgcctccccccgcgtctcttcgcagcgccaagttgtcctaagttg 182
QY	373 ctccgtctgcgggaatatcttgcgtctctaccagcgctggggggccaaagaccag 429
DB	183 ccccgctctgcagcagggctctccctctaccagcgctggggggccaaagaccag 239

[illegible][illegible]

Df		25-JAN-2001	(first entry)
Xx			
De	Eucalyptus grandis transcription factor DNA sequence #337.		
Xx			
Kw	Plant; transcription factor; gene expression; eucalyptus; pine; acacia;		
Kw	poplar; sweetgum; teak; mahogany; bzlp; G-box binding factor;		
Kw	basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;		
Kw	homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;		
Xx	type 2 Cys2His2; CCAAT box element; MYB; ss.		
Xx			
Os	Eucalyptus grandis.		
Pn	MO200053724-A2.		
Pd	14-SEP-2000.		
Xx			
Pf	09-MAR-2000; 2000WO-US06112.		
Pr	11-MAR-1999; 99US-0266513.		
Pr	18-AUG-1999; 99US-0149485.		
Xx			
Pa	(GENE-) GENESIS RES & DEV CORP LTD.		
Pa	(FLET-) FLETCHER CHALLENGE FORESTS LTD.		
Pi	Wood M, McGrath A, Shenk MA, Glenn M;		
Xx	WPI: 2000-579369/54.		
Dr			
Pt	New isolated polynucleotide encoding a plant transcription factor for		
Pt	producing a plant e.g. a woody plant, preferably eucalyptus or pine,		
Pt	having modified gene expression or modified activity of a polypeptide		
Pt	-		
Xx			
Ps	Claim 1; Page 442: 747pp; English.		
Xx			
Cc	The present invention relates to novel plant transcription factors from		
Cc	Eucalyptus grandis or Pinus radiata. The present sequence is the coding		
Cc	sequence for one such transcription factor. The transcription factor may		
Cc	be used to produce a plant having modified gene expression such as a		
Cc	woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or		
Cc	mahogany species or to modify the activity of a polypeptide in a plant.		
Cc	The transcription factors of the present invention are members from the		
Cc	following families of regulatory proteins: bZIP, bZIP family of G-box		
Cc	binding factors, basic helix-loop-helix zipper,		
Cc	homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2		
Cc	and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements		
Cc	and MYB.		
Sq			
	Sequence 420 BP; 75 A; 175 C; 115 G; 55 T; 0 other:		
	Query Match	10.8%; Score 46.2; DB 21; Length 420;	
	Best Local Similarity	51.8%; Pred. No. 0.44;	
	Matches 128; Conservative	0; Mismatches 118; Indels 1; Gaps 1	
Qy	160 accaccacacacacacacacagcgacgcgcacacgtcaacagcgcgaggaaggccg	219	
Db	50 aactcgccacacacgcgcgcgcccttcgcccacgacgcgcgcgcgcacacgttg	109	
Qy	220 cactcgccccaggggtggtgcgaagaagagcgtcgcgscgcgcgcgcgtctcagga	279	
Db	110 aaacggaagcgtccaagagcccccgcgcgcgcgcaccccaagaccagc-ccttcga	168	
Qy	280 gaacctcgcgctgtcatgtcatgtctctccgcgcgcgcgcgcacacgcgcgcgc	339	
Db	169 gtactctgcccctctgctatcatatgtctgcgcgcgcgcgcgcgcgcgcgcgc	228	
Qy	340 ggcgcgcgcgcgtccgcgtccgcgcgaagtcaagtgtctcgctctgcgcgcgcgc	399	
Db	229 caggtccacagcgtgtcatctgcacacaagccttccccacacgcgcgcgcgcgc	288	
Qy	400 ctaccagc 406		

Db	289	ccacaag	295
<hr/>			
RESULT	13		
ID	A52711		
XX	A52711 standard; cDNA; 738 BP.		
AC	A52711;		
XX			
DT	27-OCT-2000 (first entry)		
XX			
DE	Rice Mlo homologue putative coding sequence #5.		
XX			
KM	Rice; Mlo homologue; disease resistance; ss.		
XX			
OS	Oryza sativa.		
XX			
FH			
FT	Key	Location/Qualifiers	
FT	CDS	1..465	
FT		/*tag= a	
FT		/product= "Mlo homologue"	
FT		/partial	
FT		/transl_except= (pos:457..459,aa:Xaa)	
FT		/note= "Xaa=unknown"	
XX			
XX	WO200036110-A2.		
PN			
PD	22-JUN-2000.		
XX			
PF	17-DEC-1999; 99WO-US30181.		
XX			
PR	18-DEC-1998; 98US-0112737.		
XX			
PA	(DUPO) DU PONT DE NEMOURS & CO E I.		
XX			
PI	Cahoon RE, Miao G, Rafalski JA, Fang Y, Sakai H, Taramino G;		
XX			
DR	WPT: 2000-431590/37.		
P-PSDB:	B01801.		
XX			
PT	New polynucleotide encoding a Mlo homologue polypeptide, useful for		
XX	creating transgenic plants with altered levels of disease resistance -		
PS	Claim 2; Page 59; 79pp; English.		
XX			
CC	The present sequence is a putative coding sequence for a Mlo		
CC	homologue from rice. It was identified by searching a leaf cDNA library		
CC	for sequences encoding proteins similar to Mlo from Hordeum vulgare		
CC	(barley) and Arabidopsis thaliana. Mlo confers resistance to Erysiphe		
CC	graminis f. sp. hordei upon the plant, and its inactivation leads to the		
CC	priming of disease resistance even if the pathogen is not present. The		
CC	gene and protein can be used to create transgenic plants which have		
CC	increased disease resistance, as well as allowing researchers to find		
CC	other resistance-conferring genes and proteins.		
SQ	Sequence 738 BP; 155 A; 227 C; 200 G; 143 T; 13 other:		
<hr/>			
Query Match	10.8%; Score 46.2; DB 21; Length 738;		
Best Local Similarity	77.1%; Pred. NO. 0.44;		
Matches 54; Conservative	0; Mismatches 16; Indels 0; Gaps 0		
OY	146 gcagcagcagcagcacaccacacacagcagcgccacgcgtcaccagcg 205		
db			
	558 cgggcgcgagcgcgcgcacacacacacacattcaggcgccacancatnngtcaccaaca 617		
OY	206 gcgacgacgag 215		
db	618 acaacgacgag 627		
<hr/>			
RESULT	14		
CS6157			

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:49:42 ; Search time 132.33 Seconds
(without alignments)
566.044 Million cell updates/sec

Title: US-09-619-643-1

Perfect score: 429
Sequence: 1 gtaaccacagcagccaccca.....ctgggggggcaaaagaccag 429

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.2	10.3	940	2	US-08-471-717-1
2	44.2	10.3	2830	1	US-07-882-292-1
3	44.2	10.3	2830	2	US-08-331-644-1
4	44.2	10.3	2830	5	PCT-US93-04102-1
5	43.4	10.1	936	1	US-08-018-977C-4
6	43.4	10.1	2091	4	US-08-899-437-22
7	43.4	10.1	2502	4	US-08-899-437-5
8	41.4	9.7	950	3	US-08-581-148C-10
9	41.4	9.7	4257	2	US-09-690-473-1
10	41.4	9.7	4257	4	US-09-259-821A-1
11	41.4	9.7	5020	4	US-08-938-291A-3
12	41.4	9.7	12001	1	US-08-458-568A-11
13	41.2	9.6	2244	5	US-08-203-532F-1
14	41.2	9.6	2244	5	PCT-US95-01882A-1
15	40.6	9.5	252	2	US-08-332-766A-28
16	39.4	9.2	11734	6	5352575-8
17	39.4	9.2	11236	1	US-07-853-913-1
18	39.2	9.1	9960	3	US-08-822-586-46
19	38.4	9.0	2051	1	US-08-343-785-7
20	38.4	9.0	2051	2	US-08-462-221-7
21	38.4	9.0	2051	3	US-08-946-458-7
22	38.4	9.0	2111	1	US-08-343-785-1
23	38.4	9.0	2111	2	US-08-462-221-1
24	38.4	9.0	2111	3	US-08-946-458-1
25	38.2	8.9	68750	4	US-09-335-409-1
26	37.8	8.8	941	2	US-08-203-532F-3
27	37.8	8.8	941	5	PCT-US95-01882A-3

28	37.8	8.8	44377	2	US-08-804-227C-7	Sequence 7, Appl
29	37.8	8.8	44377	2	US-08-804-198-1	Sequence 1, Appl
30	37.6	8.8	1514	2	US-08-454-267-1	Sequence 1, Appl
31	37.6	8.8	1514	2	US-08-941-319-1	Sequence 1, Appl
32	37.6	8.8	1514	4	US-09-035-098-1	Sequence 1, Appl
33	37.6	8.8	3382	2	US-08-682-847-1	Sequence 1, Appl
34	37.6	8.8	3519	1	US-08-035-558-1	Sequence 1, Appl
35	37.2	8.7	7892	2	US-07-916-098A-40	Sequence 40, Appl
36	36.8	8.6	2241	2	US-08-838-219B-20	Sequence 20, Appl
37	36.8	8.6	2241	4	US-09-233-336A-20	Sequence 20, Appl
38	36.8	8.6	2241	4	US-09-233-752A-20	Sequence 20, Appl
39	36.8	8.6	2370	2	US-08-838-219B-19	Sequence 19, Appl
40	36.8	8.6	2370	4	US-09-233-336A-19	Sequence 19, Appl
41	36.8	8.6	2370	4	US-09-233-752A-19	Sequence 19, Appl
42	36.8	8.6	2403	1	US-08-471-033-30	Sequence 30, Appl
43	36.8	8.6	2403	2	US-08-471-044-30	Sequence 30, Appl
44	36.8	8.6	2403	2	US-08-463-483A-30	Sequence 30, Appl
45	36.8	8.6	2403	2	US-08-471-046A-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-08-471-717-1
Sequence 1, Application US/08471717
Patent No. 5859337
GENERAL INFORMATION:
APPLICANT: Gasser, Charles S.
APPLICANT: Lippuner, Veronica
TITLE OF INVENTION: GENES CONFERRING SALT TOLERANCE AND
TITLE OF INVENTION: THEIR USES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew
STREET: One Market Plaza, Stewart Street Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,717
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 2307E-606
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 940 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 55..738
US-08-471-717-1
Query Match 10.3%; Score 44.2; DB 2; Length 940;
Best Local Similarity 58.7%; Pred. No. 0.54;
Matches 98; Conservative 0; Mismatches 63; Indels 6; Gaps 1;

[illegible]

```

RESULT 2
US-07-882-292-1
Sequence 1, Application US/07882292
Patent No. 5324638
GENERAL INFORMATION:
APPLICANT: Tao, Wufan
APPLICANT: Lai, Esseng
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: c/o Cooper and Dunham, 30 Rockefeller
STREET: Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/882,292
FILING DATE: 19920513
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41472
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-664-0525
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: N
FEATURE:
NAME/KEY: CDS
LOCATION: 443..1882
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: 926..1855
OTHER INFORMATION: /note= "nucleotide sequence encoding DNA"
OTHER INFORMATION: Binding domain homology"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 1883..1885
OTHER INFORMATION: /note= "translation termination codon"
US-07-882-292-1

```

[illegible]

```

RESULT      3
US-08-331-644-1
: Sequence 1, Application US/08331644
: Patent No. 5976872
: GENERAL INFORMATION:
: APPLICANT: Tao, Wufan
: APPLICANT: Lai, Esseng
: TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/331,644
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/882,292
: FILING DATE: 13-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-278-0400
: TELEFAX: 212-391-0525
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2830 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cdna
: HYPOTHETICAL: N
: ANTI-SENSE: N
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 443..1882
: OTHER INFORMATION:
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 926..1255
: OTHER INFORMATION: /note= "nucleotide sequence
: OTHER INFORMATION: encoding DNA binding domain
: OTHER INFORMATION: homology"

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:03:51 ; Search time 3095.41 Seconds
(without alignments)
1210.756 Million cell updates/sec

Title: US-09-619-643-1

Perfect score: 429

Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaaaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:**

2: gb_est2:**

3: gb_est3:**

4: gb_est4:**

5: gb_est5:**

6: gb_est6:**

7: gb_est7:**

8: gb_est8:**

9: gb_est9:**

10: gb_est10:**

11: gb_est11:**

12: gb_est12:**

13: gb_est13:**

14: gb_est14:**

15: gb_est15:**

16: gb_est16:**

17: gb_est17:**

18: gb_est18:**

19: gb_est19:**

20: gb_est20:**

21: gb_est21:**

22: gb_est22:**

23: gb_est23:**

24: gb_est24:**

25: gb_est25:**

26: gb_est26:**

27: gb_est27:**

28: gb_est28:**

29: gb_est29:**

30: gb_est30:**

31: gb_est31:**

32: gb_est32:**

33: gb_est33:**

34: gb_est34:**

35: gb_est35:**

36: gb_est36:**

37: gb_est37:**

38: gb_est38:**

39: gb_est39:**

40: gb_est40:**

41: gb_est41:**

42: gb_est42:**

43: gb_est43:**

44: gb_est44:**

45: gb_est45:**

44: em_esthum10:**
45: em_esthum11:**
46: em_esthum12:**
47: em_esthum13:**
48: em_esthum14:**
49: em_esthum15:**
50: em_esthum16:**
51: em_esthum17:**
52: em_esthum18:**
53: em_esthum19:**
54: em_esthum20:**
55: em_esthum21:**
56: em_esthum22:**
57: em_esthum23:**
58: em_esthum24:**
59: em_esthum25:**
60: em_esthum26:**
61: em_esthum27:**
62: em_esthum28:**
63: em_esthum29:**
64: em_esthum30:**
65: em_esthum31:**
66: em_esthum32:**
67: em_esthum33:**
68: em_esthum34:**
69: em_esthum35:**
70: em_esthum36:**
71: em_esthum37:**
72: em_esthum38:**
73: em_esthum39:**
74: em_esthum40:**
75: em_esthum41:**
76: em_esthum42:**
77: em_esthum43:**
78: em_esthum44:**
79: em_esthum45:**
80: em_esthum46:**
81: em_esthum47:**
82: em_esthum48:**
83: em_esthum49:**
84: em_esthum50:**
85: em_esthum51:**
86: em_esthum52:**
87: em_esthum53:**
88: em_esthum54:**
89: em_esthum55:**
90: em_esthum56:**
91: em_esthum57:**
92: em_esthum58:**
93: em_esthum59:**
94: em_esthum60:**
95: em_esthum61:**
96: em_esthum62:**
97: em_esthum63:**
98: em_esthum64:**
99: em_esthum65:**
100: em_esthum66:**
101: em_esthum67:**
102: em_esthum68:**
103: em_esthum69:**
104: em_esthum70:**
105: em_esthum71:**
106: em_esthum72:**
107: em_esthum73:**
108: em_esthum74:**
109: em_esthum75:**
110: em_esthum76:**
111: em_esthum77:**
112: em_esthum78:**
113: em_esthum79:**
114: em_esthum80:**
115: em_esthum81:**
116: em_esthum82:**

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

Result No.	Score	% Match	Query Length	DB ID	Description
1	201.4	46.9	509	174	BG240203
2	78	18.2	342	164	BE229898
3	75	17.5	799	146	BF264540
4	73.2	17.1	407	148	BF429386
5	68.2	15.9	222	175	D24206
6	59	13.8	925	229	CNS0091P
7	57.8	13.5	527	119	AW720367
8	57.4	13.4	350	162	BE050406
9	55.6	13.0	535	9	AA552370
10	54.6	12.7	385	30	AA426571
11	54.6	12.7	426	29	AA408053
12	54.6	12.7	429	30	AA425774
13	52.8	12.3	139	139	BE802800
14	52.4	12.2	534	9	AA595288
15	52.2	12.2	839	229	CNS004NB
16	52.2	12.2	925	229	CNS0091P
17	52.2	12.2	935	229	CNS006KX
18	52	12.1	645	229	CNS01213
190:					em_gss_pln1:*
191:					em_gss_pln2:*
192:					em_gss_pro:*
193:					em_gss_r0d1:*
194:					em_gss_r0d2:*
195:					em_gss_r0d3:*
196:					em_gss_r0d4:*
197:					em_gss_r0d5:*
198:					em_gss_vrt1:*
199:					em_gss_vrt2:*
200:					em_gss_vrt3:*
201:					em_gss_vrt4:*
202:					em_gss_vrt5:*
203:					em_gss_vrt6:*
204:					em_gss_vrt7:*
205:					em_gss_vrt8:*
206:					em_gss_vrt9:*
207:					em_gss_vrt10:*
208:					em_gss_vrt11:*
209:					em_gss_vrt12:*
210:					em_gss_vrt13:*
211:					em_gss_vrt14:*
212:					em_gss_vrt15:*
213:					em_gss_vrt16:*
214:					em_gss_vrt17:*
215:					em_gss_vrt18:*
216:					em_gss_vrt19:*
217:					em_gss_vrt20:*
218:					em_gss_vrt21:*
219:					em_gss_vrt22:*
220:					em_gss_vrt23:*
221:					em_gss_vrt24:*
222:					em_gss_vrt25:*
223:					em_gss_vrt26:*
224:					em_gss_vrt27:*
225:					em_gss_vrt28:*
226:					em_gss_vrt29:*
227:					em_gss_vrt30:*
228:					em_gss_vrt31:*
229:					em_gss_vrt32:*
230:					em_gss_vrt33:*
231:					em_gss_vrt34:*
232:					em_gss_vrt35:*
233:					em_gss_vrt36:*
234:					em_gss_vrt37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

FEATURES	Location/Qualifiers
source	1. .509

Clones to be sequenced were prepared by mass excision."

46; Gaps 5;

RESULT

source

```
/clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
```

```

/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPRI
vector at 5' end with EcoRI and 3' end with Xho I site"
BASE COUNT      71 a      100 c      125 g      46 t
ORIGIN

Query Match
Best Local Similarity 66.4%; Pred. No. 2e-07;
Matches 142; Conservative 0; Mismatches 70; Indels 2; Gaps 2;

QY 102 cgtgacttcacatgagactcccaagcgatccatgagctgcagacgacgac 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 CCGGTGCTCATGAGAGCCCTCCACCCCGGTGCTCAAGAGAGACACACGACG 190

QY 162 caaccacacacacacagagagagagcgccacgctcacacagagagagcgccca 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 GGTGAGAGAGCGGACGCTGTGTAAGAGACAGCCGACGAGAGGAGAGGCGGACA 250

QY 222 cctgccccaggggtggtgccaagaagaagcgtcgcgcgccagcgtctcagaagagaga 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 CCGTCCCGAGGGGTGGCGAGAGCGGTCGCCGCCGACGCG-ATCGAGAGAGAGAGA 309

QY 282 accctgcgtgtgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 ACCTCGCGCTGTG-CTCTCATGCTGCGCCGCG 342

RESULT 3
BF264540 799 bp mRNA EST 17-NOV-2000
LOCUS BF264540
DEFINITION HV.CEa009L06f Hordeum vulgare seedling green leaf EST library
HV.CEa009L06f, mRNA sequence.
ACCESSION BF264540
VERSION BF264540.1 GI:11195534
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae;
Hordeum.
1 (bases 1 to 799)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
Wood,T.
Development of a genetically and physically anchored EST resource
for barley genomics
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACTCTCACTAAAGG
High quality sequence start: 6
High quality sequence stop: 559.
Location/Qualifiers
1..799
/organism="Hordeum vulgare"
/cultivar="C116155 (Mal3)"
/db_xref="taxon:4513"
/clone="HV.CEa009L06f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDNA004 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

```

```

BASE COUNT      137 a      238 c      275 g      140 t      9 others
ORIGIN

Query Match
Best Local Similarity 71.4%; Pred. No. 8.6e-07;
Matches 110; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 190 gccaccgtacacagcgagcgagcgccacctgtgccccaggggtgggccaagaagaag 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 GCACACTACACTGCGTGTAGTAGAGCGGCACGTCGTCTCAGAGATGGCCCAAGAGAA 275

QY 250 cgtcgcgcgccacgagctcacaagaagaacctgcgtctgcgtctcgtctc 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 CGATCGCGCGCGCACCG-CTCTGAGAGAGAGAACTCGCGCTCTGCTCATGCTCTGC 334

QY 310 ccgagcgagacgcccacgctgcacagcgcgcgcg 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 CCNTGGCGCGCGCGGCGGTGCGCAAGCCAAAG 368

RESULT 4
BF429386 407 bp mRNA EST 29-NOV-2000
LOCUS BF429386
DEFINITION WHE1804_F01_L02S Secale cereale anther cDNA library Secale cereale
cDNA clone WHE1804_F01_L02, mRNA sequence.
ACCESSION BF429386
VERSION BF429386.1 GI:11441427
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae;
Secale.
1 (bases 1 to 407)
Anderson,O.D., Butler,E., Chao,S., Chol,D.W., Close,T.J., Fenton
,R.D., Gustafson,J.P., Han,P.S., Hala,C.C., Kang,T., Lazo,G.R.,
Miller,R., Rausch,C.J., Ross,R., Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Another cDNA library from rye
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
Location/Qualifiers
1..407
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE1804_F01_L02"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
greenhouse. Anthers were harvested and pooled from early
meiosis to late meiosis. The tissue, total RNA, and
poly(A) RNA were prepared (Butler, Ross and Gustafson) at
University of Missouri, Columbia. A cDNA library was
made, and the cDNA clones were in vivo excised to give
pBluescript phagemids in the T7 Close lab (Chol, Close,
Fenton) at the University of California, Riverside.
Plasmid DNA preparations and DNA sequencing were performed
in the OD Anderson lab (all other authors)."

```

```

BASE COUNT      57 a      155 c      151 g      44 t

```


OY	185	gagcgccacccgtcaaccagcgcgaagcggcgcccacactgccccgggtgggccaga	244
Dd	729	CSSSVSCSSVASMCCSSSSSSASAKSSSSSSSABACASCCTTWSGCSCTISASHMAA	670
OY	245	aagaagcgtctgcggcgccacgcgtlctcagaaggagaaacctgcgcgtgtcgtctcatg	304
Dd	669	RSSSSSSSSCASSMSASASSASSSSSSSSSGCGAGCAGBSMSSGGGSGGSVASG	610
OY	305	cctcccgcgagcagcacgcgcgttccagcgcgcgccgcggttcccggtccgcgcg	364
Dd	609	MSSSVSSSGRRSSGCGGGGVSGBSSSSGSGSGVCSCSCRCSCSSAAAAA	550
OY	365	ttcaggtgtcgcgtctg	381
Dd	549	ASCVAAACGMCGKSKS	533
RESULT	7		
LOCUS	AW720367	527 bp	mRNA EST 19-APR-2000
DEFINITION	LJNEST1g11r Lotus japonicus nodule library 5 and 7 week-old Lotus japonicus cDNA 5', mRNA sequence.		
ACCESSION	AW720367		
VERSION	AW720367.1	GI:7614906	
KEYWORDS	EST.		
ORGANISM	Lotus japonicus.		
REFERENCE	Lotus japonicus. Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus. 1 (bases 1 to 527) Colebatch,G., Freund,S., Trevaaskis,B and Udvardi,M. Lotus japonicus root nodule ESTs: tools for functional genomics unpublished (2000) Contact: Udvardi MK Molecular Plant Nutrition Max Planck Institute of Molecular Plant Physiology Am Muehlenberg 1, 14476 Golm, Germany Fax: 49 331 567 8250 Email: udvardi@mpimp-golm.mpg.de Seq primer: T7 High quality sequence stop: 527. Location/Qualifiers 1..527 /organism="Lotus japonicus" /cultivar="Gifu (B-129)" /db_xref="taxon:34305" /clone_11b="Lotus japonicus nodule library 5 and 7 week-old" /dev_stage="5 and 7 week-old plants" /note="Organ: Nodule; Vector: pSPORFL; Site_1: SalI; Site_2: NotI; The library was prepared using mRNA extracted from nodules of 5 and 7 week-old Lotus plants. Nodules were induced by, and contained Mesorhizobium strain R7A."		
FEATURES	source		
BASE COUNT	114 a	196 C	110 G 107 t
ORIGIN			
Query Match	13.5%	Score 57.8;	DB 119; Length 527;
Best Local Similarity	51.9%;	Pred. No. 0.0045;	
Matches 201; Conservative	0;	Mismatches 157;	Indels 29; Gaps 2
OY	72	ctccacgcgaacgttagagcttgccgtgctgcgtgatctcatcagagctaactcaacgcgt	131
Dd	2	CTCTTTTCCAATTTTCAGATTTCACATTCTCTCTCTCATGTTTCAGAAATCAGATTCACA	61
OY	132	catcataggagctgcgaacgaacgaacgaacacacacacacacacacagcagcagcgcg	191
Dd	62	AATTAATTAAAGGAAGGCATGGAACCCCTCAACTCACACCACACACCGCCCATTCCTT	121
OY	192	cacgcgtacacagcggcgaagcagggcgcccactctgccccggggtggccaagaagaagc	251

Dd	122	CACGTTCAACGACACCAACCCTCCGTACCCCGGGAGCGCTGGGCANAAOGGAACG	181
Oy	232	ctcgcgagccacagt-----ctcagaagagagaacctcgctgtgttcgtctcatgt	306
Dd	182	CTTCAAAGATCCTGTCATGCTCCGAGGAMAGAATCACCTGCTCTCTCATCATGTCT	241
Oy	307	ctccgcggagacgcgcacaccgctcgaagcgcgccgcggttccgctcgcgcgaag--	364
Dd	242	GCTGTGCGGGCGGCCGACACCATCTCCACCCCTCCGCTGCAACCGGCAACCGCTCC	301
Oy	365	-----ttcagtgctcgcgtctctcgtcggaatctctgctccta	402
Dd	302	TTCGGTTCCTCCAGGCTCAGCTACAGTGCTCCGTTGTCGACAAAGGCTTCCCTTCTTA	361
Oy	403	ccaagcgctggggggccaagaaccag	429
Dd	362	CCAACGCTGCGGAGACACAAGGCCAG	388
RESULT	8		
LOCUS	BE050406	350 bp	mRNA EST
DEFINITION	za66c01.b50 Maize Glume cDNAs Library zea mays cdna clone za66c01		
ACCESSION	BE050406		
VERSION	BE050406.1	GI:8367461	
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Mangoliopsida; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 350) O'Shaughnessy,A.L., Habermann,K., de la Bastide,M., Huang,E.N., Nascimento,L.U., Schults,K., Matero,A., Swaby,I., See,L.-H., Preston, R.R., Rodriguez,M.A., Sheth,R.S., Shekher,M., Spiegel,L.A., Vili, 'M.D., Dechla,N.N. and McCombie,W.R. Expressed sequence tags from Zea mays (maize) Unpublished (2000) Contact: W. Richard McCombie Lila Annenberg Hazen Genome Sequencing Center Cold Spring Harbor Laboratory PO Box 100, Cold Spring Harbor, NY 11724, USA Tel.: 516 367 8884 Fax: 516 367 8874 Email: mccombie@cshl.org Plate: za66 row: c column: 01 Seq primer: -40M13ForUnitv High quality sequence stop: 350. Location/Qualifiers .350 /organism="Zea mays" /db_xref="taxon:4577" /clone="za66c01" /clone_lib="Maize Glume cDNAs Library" /note="Vector: Lambda zap II (Stratagene); Site_1: XhoI; Site_2: EcoRI; Resistance: Ampicillin Autoexcision: phbscript SK (+/-) Titer: 7 x 10e-9 pfu/ml (as of 9/28/94)"		
FEATURES	source		
BASE COUNT	71 a 118 c 103 g 54 t 4 others		
ORIGIN			
Query Match	13.4%; Score 57.4; DB 162; Length 350;		
Best Local Similarity	86.2%; Pred. No. 0.0055;		
Matches	75; Conservative 0; Mismatches 11; Indels 1; Gaps 1.		
Oy	226	ccccaggaggtgggccaagaagagcgtcgcgcgccagcgtctcagaaggagaaacct	285
Dd	265	CCAGGTCGTGGGCCAAGAGGAGGCTCGGCCGCCACAGC-ATTGGAGAGAGAACACT	323
Oy	286	cggcgtgtgctgtcattgctctccc	312

REFERENCE 1 (bases 1 to 839)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers

FEATURES
 source 1..839
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCT-98"
 /clone="BACR10E16"
 /note="end : TET3"

BASE COUNT 285 a 67 c 77 g 27 t 383 others
 ORIGIN

Query Match 12.2% Score 52.2; DB 229; Length 839;
 Best Local Similarity 18.8% Pred. No. 0.073;
 Matches 63; Conservative 133; Mismatches 139; Indels 0; Gaps 0;

```

OY 39 gccagaagagatcgctcgcgcgtccatccacgaacgtagagcttgcgtg 98
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 504 SSSASGSCAGSSSSAGSGGAGSGAGSGGSGGSGGSGGSGGSGGSGGSA 563
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 99 tgcgtgacttcattgagctactccacgcgttcattcattgagctgacgacgac 158
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 564 CSCASASASSSSSSAGSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 623
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 159 gaccacacacacacacacacgacgacgacgacgacgacgacgacgacgac 218
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 624 ARSGARGMGAGAGSAGRAVSAASAAASASASASASASASASASASASAS 683
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 219 ccactgccccagagggtgagccaagaagaagcgtccgcgcgcgtctcagagagg 278
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 684 CASVCASSASGSCGSCSSMCASVSGAASSGASGCGCCSSSSGCSGCSSSSMSCS 743
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 279 agaactcgcgtctgctgctcattcctccgcgcgcgcgcgcgcgcgcgcgc 338
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 744 SSCSSSGCCSVSCSCSVSCSCBSGSCBSGSCBSGSCBSGSCBSGSCBSGSC 803
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 339 cggcgccgcgcgtcccgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 373
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 804 MCAMSSASSASSSSSSAGSGSSGSAVSSCGAVVGSS 838
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 8, 2001, 05:34:21
 Job time: 9030 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:45:07 ; Search time 2791.02 Seconds
(without alignments)
2182.308 Million cell updates/sec

Title: US-09-619-643-2
Perfect score: 413
Sequence: 1 agtcgttaatcattgca.....ccccaggtcgaaccgag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rnd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rnd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	61	14.8 167405	13 AP002483	AP002483 Oryza sat
2	40.6	9.8 110665	91 HS696P19	AL035588 Human DNA
3	38.4	9.3 110000	84 HSY237C10.0	AL031601 Homo sap1
4	37	9.0 1134	94 MUSCS1806	M64849 Mouse plate
5	37	9.0 1134	94 MUSCS1806	M84453 Mus musculu
6	36.2	8.8 9625	65 AC017264	AC017264 Drosophill
7	36.2	8.8 113224	60 AC007588	AC007588 Drosophill
8	36.2	8.8 240495	4 AE003815	AE003815 Drosophill
9	35.2	8.5 1922	93 MMU291498	AJ291498 Macaca mu
10	35.2	8.5 1922	93 MMU291499	AJ291499 Macaca mu
11	35.2	8.5 89690	12 AC079374	AC079374 Arabidops

C 12	35	8.5	104246	77	AC004785	AC004785 Arabidops
C 13	35.2	8.5	105863	12	AC004133	AC004133 Genomict s
C 14	35	8.5	1852	3	SACTYA	AtJ223796 Stigmatel
C 15	35	8.5	1922	93	MMU242441	AtJ24441 Macaca mu
C 16	34.6	8.4	145825	60	AC005729	AC005729 Homo sapi
C 17	34.2	8.3	186974	83	AP003071	AP003071 Homo sapi
C 18	34.2	8.3	194781	66	AC009124	AC009124 Homo sapi
C 19	34	8.2	125188	80	AL355884	AL355884 Homo sap
C 20	34	8.2	135490	64	AC015478	AC015478 Homo sap
C 21	34	8.2	196840	68	AC023255	AC023255 Homo sap
C 22	33.8	8.2	226772	66	AC021061	AC021061 Mus muscu
C 23	33.6	8.1	438759	86	AC006133	AC006133 Homo sapi
C 24	33.6	8.1	123814	85	AC005393	AC005393 Homo sapi
C 25	33.6	8.1	166809	81	AL391318	AL391318 Homo sap
C 26	33.4	8.1	10822	1	AE004905	AE004905 Pseudom
C 27	33.2	8.0	160558	70	AC025832	AC025832 Homo sapi
C 28	33.2	8.0	163764	92	HSDB46P14	AL109800 Human
C 29	33.2	8.0	167691	91	HS406A7	AL023284 Human DNM
C 30	33	8.0	2396	94	MUSDC47	D26091 Mus musculi
C 31	33	8.0	64745	61	AC008754	AC008754 Homo sapi
C 32	33	8.0	171509	61	AC009870	AC009870 Homo sapi
C 33	33	8.0	176120	75	AC073548	AC073548 Homo sapi
C 34	32.8	7.9	42167	85	AC000030	AC000030 Homo sapi
C 35	32.8	7.9	43847	85	AC000036	AC000036 Homo sapi
C 36	32.6	7.9	172931	86	AC006581	AC006581 Homo sapi
C 37	32.6	7.9	203918	68	AC024093	AC024093 Homo sapi
C 38	32.4	7.8	165659	88	AF003529	AF003529 Homo sapi
C 39	32.4	7.8	171188	60	AC005025	AC005025 Homo sapi
C 40	32.4	7.8	172983	85	AC002454	AC002454 Human BAC
C 41	32.2	7.8	1434	8	GGA23592	AtJ237599 Gallus ga
C 42	32.2	7.8	42537	77	AC0084302	AC0084302 Homo sapi
C 43	32.2	7.8	97982	90	AP000687	AP000687 Homo sapi
C 44	32.2	7.8	171703	90	AP000688	AP000688 Homo sapi
C 45	32.2	7.8	340000	90	AP001724	AP001724 Homo sapi

IGNMENTS

AC0841785	Arbidiops
AC0041133	Genomic s
AJ223796	Stigmatel
AF1424441	Macaca mu
AC055279	Homo sapi
AP003011	Homo sapi
AC019174	Homo sapi
AL155884	Homo sapi
AC015478	Homo sapi
AC023255	Homo sapi
AC021051	Mus muscu
AC061133	Homo sapi
AC005393	Homo sapi
AL191318	Homo sapi
AE004905	Pseudomon
AC025832	Homo sapi
AL109800	Human DN
AF123284	Human DN
D820121	Mus muscul
AC0081754	Homo sapi
AC009870	Homo sapi
AC073548	Homo sapi
AC000050	Homo sapi
AC000036	Homo sapi
AC006581	Homo sapi
AC024053	Homo sapi
AF003529	Homo sapi
AC005425	Homo sapi
AC0020454	Human BAC
AJ237599	Gallus ga
AC084302	Homo sapi
AP000667	Homo sapi
AP000668	Homo sapi
AP001724	Homo sapi

FEATURES

gene
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0019D06"
7267..7644
/gene="P0019D06.1"
7267..7644
/gene="P0019D06.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB16446.1"
/db_xref="GI:10798815"
/translation="MEKEKHGQRRRLRSGRPALFTDAIHAGRRHDAQPTSTQ
VNPDEGRHLPRLRCRRTBASLWTEELDKSAATDADKAGGSGSMITSTIK
VKNDDEKLAAMASAMPTRSP"
complement(join(8744..8754,9205..9316,9401..9556,
9860..9958,10073..10147,10266..10334,10501..10699,
11076..11181,11702..12110))
/gene="P0019D06.2"
complement(join(8744..8754,9205..9316,9401..9556,
9860..9958,10073..10147,10266..10334,10501..10699,
11076..11181,11702..12110))
/gene="P0019D06.2"
/note="contains EST AU100943(C60656)"
/codon_start=1
/product="putative AAA-type ATPase"
/protein_id="BAB16447.1"
/db_xref="GI:10798816"
/translation="MYSNFKQALIEYKQAOEDNGCNYYKAPFLYNNALLEYFTHLK
YEQPKINDATAAKFTEYLRRAEIEIRVLDSHIGGGGDSQEQAKPRGMRSAIVAAK
KPSKWSDVAGLESAKALQEAALPIKPFHFGKRSPMKAFLLVGPFGKSYLAB
AVAEVDSFTESSISDLSVKMGSEKLVANLEQMARENAPSIIFIDEISLGCQSGY
CENENASPRKIKTELLVOMOGFDSNDKRVLIATNMPHYLDQMRRRPFKCIYIP
DKRKDKPKRIKIGDTPHSLTEGDFSLAVOTEEGSDIAYCYKDALFOVRYKTODA
KFKKADDDTPVTESEQSGPSIQTTMQLASGLAAKILPPIKIDPELVQRPPT
VSKDLVYKEFTFEFTRL"
join(13646..13747,14696..14806)
/gene="P0019D06.3"
join(13646..13747,14696..14806)
/gene="P0019D06.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB16448.1"
/db_xref="GI:10798817"
/translation="MWSMCPQKRLFTDGTSGQHREFGAIDNLVQASCEGHMINTR
IHTGSETIFDSRAEIHNSPSSNKRND"
complement(join(15584..16217,16229..17164))
/gene="P0019D06.4"
complement(join(15584..16217,16229..17164))
/gene="P0019D06.4"
/note="probably inactive due to frameshift in CDS
pseudogene, hypothetical protein
similar to Oryza sativa chromosome 1, P0031E09.21"
/codon_start=1
/pseudo
join(17888..18090,18725..18854)
/gene="P0019D06.5"
join(17888..18090,18725..18854)

This sequence is the entire insert of clone 696P19. This sequence has been finished according to sequencing map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

696P19 is from the library RCI-4 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2>.

FEATURES

source

1. .110665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p12.3-21.2"
/clone_lib="RCI-4"
/clone="RP4-696P19"
36. .139
/note="MIR repeat: matches 19. .139 of consensus"
394. .638
/note="MIR repeat: matches 14. .259 of consensus"
1489. .1610
/note="MER5B repeat: matches 11. .154 of consensus"
1625. .1801
/note="MER5A repeat: matches 1. .189 of consensus"
2672. .2775
/note="L2 repeat: matches 2645. .2748 of consensus"
3417. .3530
/note="L2 repeat: matches 2581. .2700 of consensus"
3811. .3974
/note="LIM3 repeat: matches -184. .-18 of consensus"
4007. .4307
/note="Aluub repeat: matches 1. .299 of consensus"
4887. .4977
/note="LIME3A repeat: matches 6074. .6162 of consensus"
4978. .5267
/note="AluSg repeat: matches 1. .293 of consensus"
5268. .5413
/note="LIME3A repeat: matches 5921. .6074 of consensus"
5443. .5576
/note="LIME3A repeat: matches 5815. .5951 of consensus"
5579. .5753
/note="AluSg/x repeat: matches 137. .312 of consensus"
5791. .5964
/note="LIMC/D repeat: matches 5394. .5568 of consensus"
6376. .6537
/note="MER20 repeat: matches 1. .218 of consensus"
6560. .6691
/note="MIR repeat: matches 62. .184 of consensus"
8277. .8372
/note="L2 copies 8 mer tctgtgtgt 76% conserved"
8277. .8374
/note="L49 copies 2 mer tg 77% conserved"
8677. .8885
/note="MIR repeat: matches 12. .240 of consensus"
9005. .9194
/note="MTR repeat: matches 65. .262 of consensus"
9437. .9736
/note="AluSg repeat: matches 1. .297 of consensus"
9902. .10079
/note="MIR repeat: matches 79. .262 of consensus"
10312. .10349
/note="L19 copies 2 mer gt 97% conserved"

repeat_region 10312. .10351
/note="5 copies 8 mer gtgtgtgt 95% conserved"
repeat_region 10788. .10902
/note="Charlie3 repeat: matches 2545. .2651 of consensus"
repeat_region 11100. .11149
/note="MIR repeat: matches 89. .139 of consensus"
repeat_region 11683. .11877
/note="MIR repeat: matches 61. .251 of consensus"
repeat_region 11880. .12078
/note="MIR repeat: matches 19. .250 of consensus"
repeat_region 12100. .12231
/note="MIR repeat: matches 68. .193 of consensus"
repeat_region 12269. .12803
/note="MER1A repeat: matches 1. .527 of consensus"
12836. .13038
/note="match: GSS: Em:AQ609073 Em:AQ563754"
13049. .13207
/note="L1MA7 repeat: matches 5990. .6114 of consensus"
13208. .13517
/note="Aluub repeat: matches 1. .310 of consensus"
13518. .13672
/note="L1MA7 repeat: matches 6114. .6266 of consensus"
13804. .13931
/note="MIR repeat: matches 21. .145 of consensus"
14160. .14295
/note="MIR repeat: matches 77. .211 of consensus"
15445. .15629
/note="MIR repeat: matches 33. .247 of consensus"
15842. .15969
/note="MIR repeat: matches 48. .178 of consensus"
16025. .16328
/note="AluSx repeat: matches 1. .304 of consensus"
17104. .17159
/note="L2 repeat: matches 2692. .2750 of consensus"
17469. .17628
/note="80 copies 2 mer ca 80% conserved"
17475. .17626
/note="L19 copies 8 mer cacacaca 80% conserved"
19195. .19312
/note="MIR repeat: matches 28. .149 of consensus"
19431. .19496
/note="2 copies 33 mer 94% conserved"
20147. .20472
/note="match: GSS: Em:AQ165690"
20473. .20588
/note="MIR repeat: matches 105. .229 of consensus"
20703. .20811
/note="L2 repeat: matches 2577. .2697 of consensus"
21659. .21933
/note="Aluub repeat: matches 32. .304 of consensus"
22065. .22144
/note="match: STS G08578"
22146. .22241
/note="L10 copies 8 mer atccatcc 74% conserved"
22146. .22241
/note="L12 copies 8 mer atccatcc 80% conserved"
22264. .22326
/note="L2 repeat: matches 2688. .2750 of consensus"
22278. .22489
/note="MIR repeat: matches 19. .254 of consensus"
23025. .23190
/note="L1R16C repeat: matches 191. .354 of consensus"
23025. .23190
/note="match: GSS: Em:B75626"
26621. .26698
/note="L39 copies 2 mer gt 65% conserved"
27047. .27142
/note="MER5A repeat: matches 55. .175 of consensus"
27258. .27464
/note="MIR repeat: matches 9. .217 of consensus"
27780. .27967
/note="MIR repeat: matches 60. .245 of consensus"
28791. .28929
repeat_region

```

repeat_region /note="MIR repeat: matches 33. 170 of consensus"
29689..29751
/note="MIR repeat: matches 47. 98 of consensus"
repeat_region /note="AluX repeat: matches 1. 302 of consensus"
29752..30053
/note="AluX repeat: matches 1. 302 of consensus"
repeat_region /note="MIR repeat: matches 98. 129 of consensus"
30054..30082
/note="MIR repeat: matches 1. 310 of consensus"
repeat_region /note="AluX repeat: matches 1. 310 of consensus"
31222..31531
/note="AluX repeat: matches 1. 310 of consensus"
repeat_region /note="LIME3 repeat: matches 6017. 6153 of consensus"
31713..31848
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
33446..33658
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
34084..34129
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
34286..34590
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
34614..34997
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
34998..35305
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
35306..35472
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
36075..36310
/note="LIME3 repeat: matches 34. 260 of consensus"
repeat_region /note="LIME3 repeat: matches 34. 260 of consensus"
36498..36655
/note="LIME3 repeat: matches 34. 260 of consensus"

Query Match 9.88; Score 40.6; DB 91; Length 110665;
Best Local Similarity 50.88; Pired. No. 0.077;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Db 50 cctccacaattatgagaccacatgaagaatcctccttgctatgctccgaa 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27809 CCTCGGCGAAGTTTCAAGCCCTCTGTCGACTCACTTCCCTCTGTAACAGGGATA 27868
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 110 gctaacccctcaatacagcgctgaagcctcagcagatactacatgatatcataag 169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27869 TGTAAAGCCCTCATCATGTGCGATGAGAGCTGTGATGATTAATCAGTAACACATAT 27928
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 170 gacactgctgcagatcagcgcaggggtggcgtgcacgctgctagtaagaagcgtct 229
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27929 CAATGCGAGAACAGTCCCTGTCGAGAGTACGACTGTCATAGTACTAATAGCCCTCT 27988
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 230 ggcctgcccagc 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27989 TGGTTGCAACC 27999
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
HSY237C10_0
WPCOMMENT
Sequence split into 5 fragments LOCUS HSY237C10 Accession AL031601
Fragment Name Begin End
HSY237C10_0 1 110000
HSY237C10_1 100001 210000
HSY237C10_2 200001 310000
HSY237C10_3 300001 410000
HSY237C10_4 400001 464225
LOCUS HSY237C10 464225 bp DNA HTG 22-NOV-1999
DEFINITION Homo sapiens chromosome 10 clone XX-Y237C10, *** SEQUENCING IN
PROGRESS ***, 34 unordered pieces.
ACCESSION AL031601
VERSION AL031601.2 GI:6981826
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 464225)
AUTHORS McMurray,A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Wellcome Trust Genome Campus, Hinxton,

```

COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquer@anger.ac.uk Clone requests: clonerequest@anger.ac.uk

On Feb 16, 2000 this sequence version replaced g1:4469255. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E. coli, yeast, vector, phage etc. Order of segments is not known. 800 n's separate segments. Unfinished: Y237C10 Contig_ID: 01233 acc=AL031601 Length: 1442 bp Unfinished: Y237C10 Contig_ID: 04833 acc=AL031601 Length: 1494 bp Unfinished: Y237C10 Contig_ID: 05043 acc=AL031601 Length: 1104 bp Unfinished: Y237C10 Contig_ID: 01429 acc=AL031601 Length: 1761 bp Unfinished: Y237C10 Contig_ID: 04844 acc=AL031601 Length: 1093 bp Unfinished: Y237C10 Contig_ID: 05117 acc=AL031601 Length: 1024 bp Unfinished: Y237C10 Contig_ID: 04903 acc=AL031601 Length: 1372 bp Unfinished: Y237C10 Contig_ID: 03908 acc=AL031601 Length: 1083 bp Unfinished: Y237C10 Contig_ID: 05079 acc=AL031601 Length: 1233 bp Unfinished: Y237C10 Contig_ID: 05119 acc=AL031601 Length: 1441 bp Unfinished: Y237C10 Contig_ID: 04755 acc=AL031601 Length: 1502 bp Unfinished: Y237C10 Contig_ID: 05097 acc=AL031601 Length: 2794 bp Unfinished: Y237C10 Contig_ID: 04816 acc=AL031601 Length: 1682 bp Unfinished: Y237C10 Contig_ID: 04781 acc=AL031601 Length: 1033 bp Unfinished: Y237C10 Contig_ID: 04774 acc=AL031601 Length: 2731 bp Unfinished: Y237C10 Contig_ID: 04780 acc=AL031601 Length: 2211 bp Unfinished: Y237C10 Contig_ID: 05052 acc=AL031601 Length: 1373 bp Unfinished: Y237C10 Contig_ID: 04997 acc=AL031601 Length: 1080 bp Unfinished: Y237C10 Contig_ID: 05056 acc=AL031601 Length: 1086 bp Unfinished: Y237C10 Contig_ID: 04934 acc=AL031601 Length: 2156 bp Unfinished: Y237C10 Contig_ID: 04957 acc=AL031601 Length: 1224 bp Unfinished: Y237C10 Contig_ID: 04739 acc=AL031601 Length: 2005 bp Unfinished: Y237C10 Contig_ID: 05031 acc=AL031601 Length: 1031 bp Unfinished: Y237C10 Contig_ID: 04947 acc=AL031601 Length: 2763 bp Unfinished: Y237C10 Contig_ID: 01067 acc=AL031601 Length: 387198 bp Unfinished: Y237C10 Contig_ID: 05024 acc=AL031601 Length: 1784 bp Unfinished: Y237C10 Contig_ID: 04820 acc=AL031601 Length: 1323 bp Unfinished: Y237C10 Contig_ID: 05006 acc=AL031601 Length: 2659 bp Unfinished: Y237C10 Contig_ID: 04949 acc=AL031601 Length: 1173 bp Unfinished: Y237C10 Contig_ID: 04801 acc=AL031601 Length: 1167 bp Unfinished: Y237C10 Contig_ID: 04863 acc=AL031601 Length: 1499 bp Unfinished: Y237C10 Contig_ID: 05087 acc=AL031601 Length: 1203 bp Unfinished: Y237C10 Contig_ID: 05012 acc=AL031601 Length: 1051 bp.

* NOTE: This is a 'working draft' sequence. It currently consists of 34 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1442: contig of 1442 bp in length

1443 2242: gap of 800 bp

2243 3736: contig of 1494 bp in length

3737 4536: gap of 800 bp

4537 5640: contig of 1104 bp in length

5641 6440: gap of 800 bp

6441 8201: contig of 1761 bp in length

8202 9001: gap of 800 bp

9002 10094: contig of 1093 bp in length

10095 10894: gap of 800 bp

10895 11918: contig of 1024 bp in length

11919 12718: gap of 800 bp

12719 14090: contig of 1372 bp in length

14091 14890: gap of 800 bp

14891 15973: contig of 1083 bp in length

15974 16773: gap of 800 bp

16774 18006: contig of 1233 bp in length

18007 18806: gap of 800 bp

```
* 18807 20247: contig of 1441 bp in length
* 20248 21047: gap of 800 bp
* 21048 22549: contig of 1502 bp in length
* 22550 23349: gap of 800 bp
* 23350 26143: contig of 2794 bp in length
* 26144 26943: gap of 800 bp
* 26944 28625: contig of 1682 bp in length
* 28626 29425: gap of 800 bp
* 29426 30475: contig of 1050 bp in length
* 30476 31275: gap of 800 bp
* 31276 32308: contig of 1033 bp in length
* 32309 33108: gap of 800 bp
* 33109 33839: contig of 731 bp in length
* 33840 36639: gap of 800 bp
* 36640 38850: contig of 2211 bp in length
* 38851 39650: gap of 800 bp
* 39651 41023: contig of 1373 bp in length
* 41024 41823: gap of 800 bp
* 41824 42903: contig of 1080 bp in length
* 42904 43703: gap of 800 bp
* 43704 44789: contig of 1086 bp in length
* 44790 45589: gap of 800 bp
* 45590 47745: contig of 2156 bp in length
* 47746 48545: gap of 800 bp
* 48546 49769: contig of 1224 bp in length
* 49770 50569: gap of 800 bp
* 50570 52574: contig of 2005 bp in length
* 52575 53374: gap of 800 bp
* 53375 54405: contig of 1031 bp in length
* 54406 55205: gap of 800 bp
* 55206 57968: contig of 2763 bp in length
* 57969 58768: gap of 800 bp
* 58769 445966: contig of 387198 bp in length
* 445967 446766: gap of 800 bp
* 446767 448550: contig of 1784 bp in length
* 448551 449350: gap of 800 bp
* 449351 450673: contig of 1323 bp in length
* 450674 451473: gap of 800 bp
* 451474 454132: contig of 2659 bp in length
* 454133 454932: gap of 800 bp
* 454933 456105: contig of 1173 bp in length
* 456106 456905: gap of 800 bp
* 456906 458072: contig of 1167 bp in length
* 458073 458872: gap of 800 bp
* 458873 460371: contig of 1459 bp in length
* 460372 461171: gap of 800 bp
* 461172 462374: contig of 1203 bp in length
* 462375 463174: gap of 800 bp
* 463175 464225: contig of 1051 bp in length.
```

```
FEATURES
Source
1. 464225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="XX-Y237C10"

BASE COUNT 132917 a 88171 c 85912 g 130770 t 26455 others
ORIGIN
```

```
Query Match 9.3%: Score 38.4; DB 84; Length 110000;
Best Local Similarity 52.5%: Pred. No. 0.39; Mismatches 76; Indels 0; Gaps 0;
Matches 84; Conservative 0;

185 catggcgacgggtgggctgcacacgtgctagtagacagcgtctgctgacgaagaact 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7756 CATGCATGCGGCGTGTCTCCGGAAGCCTGTGATGACGAGCCCTTCGCGTACGATATTGCG 7815
245 gatcatgcatagggcgaacacgtcgaagcctctatgtgtcaatcaatccatccgtlagccg 304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7816 CATCATGCAAGACGAGACCATCTGCGCATATACCGTTCTACCCATCATGCGGCGTACT 7875
305 gtgacgaaggtccggttcgtgtagcttgaggcgacgaact 344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 7876 GGAGCGAAGCGGCGATGACCGGAGACCTGCTGAGACTGT 7915
RESULT 4
MUSCIS06 1134 bp DNA ROD 27-APR-1993
LOCUS Mouse platelet-derived growth factor B chain (c-sis) gene, exons 6
DEFINITION and 7.
ACCESSION M64849.1 GI:192818
VERSION M64849.1
KEYWORDS c-sis proto-oncogene; platelet-derived growth factor;
platelet-derived growth factor B chain; proto-oncogene.
SEGMENT 6 of 6
SOURCE Mouse DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1134)
AUTHORS Bonchiron,D.T., Sultan,P. and Collins,T.
TITLE Structure of the murine c-sis proto-oncogene (Sis, PDGFR) encoding
the B chain of platelet-derived growth factor
JOURNAL Genomics 10, 287-292 (1991)
MEDLINE 91257844
```

```
FEATURES
Source
1. 1134
/organism="Mus musculus"
/db_xref="taxon:10090"
join(M64844.1:2070..3393,M64845.1:1..601,M64846.1:1..928,
M64847.1:1..223,M64848.1:1..311,1..1134)
/gene="c-sis"
join(M64844.1:3046..3108,M64845.1:302..398,
M64846.1:824..913,M64847.1:16..221,M64848.1:141..285,
10..134)
/gene="c-sis"
/codon_start=1
/product="platelet-derived growth factor B chain"
/protein_id="AA37485.1"
/db_xref="GI:192820"
/translation="MNRCAWLFPLCYLRLVSAEGDIPPELYEMLSHSIRSFIDL
OQLHRDSDVEGALDLMNTRAHSGVLESSRRRLGSLAAEPVIVAECKRTRE
VFOISRLIDRTNANFLVPCVEVORSGCCNNRNVCCRAVOVMREVQVRIETVR
KKPIRKATVLEHLACKCEITVPRPVTRSPGSRBQRAKTPQARVITVRIRRP
PKGKRRKRHHDKAALKETLGA"
order(M64848.1:286..311,<1..9)
/gene="c-sis"
/number=5
/number=5
/gene="c-sis"
/number=6
/number=6
/gene="c-sis"
/number=6
/number=6
/gene="c-sis"
/number=7
/number=7

exon
intron
intron
exon

BASE COUNT 215 a 365 c 292 g 262 t
ORIGIN
```

```
Query Match 9.0%: Score 37; DB 94; Length 1134;
Best Local Similarity 57.3%: Pred. No. 0.79; Mismatches 50; Indels 0; Gaps 0;
Matches 67; Conservative 0;

176 tgcctcatcatgagcgacgggtgggctgcacacgtgctagtagacgaagcgtctggtgg 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
560 TGTGTCTTCCCTTCACCCGCGTGTGCCGCGCACCTCCACACGACGAGATCCCTACGCGCC 619
236 ccagcaactgatactgataagggcgaactgcgaagcctctatgtgtcaatcaatcca 292
```


AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
Celniker,S.E., Agbayani,A., Arcalena,T.T., Baxter,E., Blazef,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Man,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.	Sequencing of <i>Drosophila melanogaster</i> Unpublished			
2 (bases 1 to 113224)				
Celniker,S.E., Agbayani,A., Arcalena,T.T., Baxter,E., Blazef,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.	Direct Submission			
Submitted (19-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA				
On Feb 11, 2000 this sequence version replaced gi:5629949.				
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (http://www.fruitfly.org/sequence/) or send email to bdg@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases.				
* NOTE: This is a 'working draft' sequence. It currently consists of 73 contigs. The true order of the pieces				
* is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will be preserved.				
1	723: contig of 723 bp in length			
724	803: gap of unknown length			
804	1569: contig of 766 bp in length			
1570	1649: gap of unknown length			
1650	2257: contig of 608 bp in length			
2258	2337: gap of unknown length			
2338	3625: contig of 1288 bp in length			
3626	3705: gap of unknown length			
3706	4738: contig of 1033 bp in length			
4739	4818: gap of unknown length			
4819	5758: contig of 940 bp in length			
5759	5839: gap of unknown length			
5839	7279: contig of 1441 bp in length			
7280	7359: gap of unknown length			
7360	8002: contig of 643 bp in length			
8003	8082: gap of unknown length			
8083	8782: contig of 700 bp in length			
8783	8862: gap of unknown length			
8863	10036: contig of 1174 bp in length			
10037	10116: gap of unknown length			
10117	10737: contig of 621 bp in length			
10738	10817: gap of unknown length			
10818	11819: contig of 1002 bp in length			
11820	11899: gap of unknown length			
11900	12843: contig of 1044 bp in length			
12944	13023: gap of unknown length			
13024	13900: contig of 877 bp in length			
13901	13980: gap of unknown length			
13981	15154: contig of 1174 bp in length			
15155	15234: gap of unknown length			
15235	15235: contig of 764 bp in length			
15999	16078: gap of unknown length			
16079	17425: contig of 1347 bp in length			
17426	17505: gap of unknown length			
17506	18832: contig of 1327 bp in length			
18833	18912: gap of unknown length			

18913	20090: contig of 1178 bp in length
20091	20170: gap of unknown length
20171	20919: contig of 749 bp in length
20920	20999: gap of unknown length
21000	22096: contig of 1097 bp in length
22097	22176: gap of unknown length
22177	23042: contig of 866 bp in length
23043	23122: gap of unknown length
23123	24265: contig of 1143 bp in length
24266	24345: gap of unknown length
24346	25208: contig of 863 bp in length
25209	25288: gap of unknown length
25289	27008: contig of 1720 bp in length
27009	27088: gap of unknown length
27089	28370: contig of 1282 bp in length
28371	28450: gap of unknown length
28451	29117: contig of 1267 bp in length
29118	29797: gap of unknown length
29798	31035: contig of 1238 bp in length
31036	31115: gap of unknown length
31116	32651: contig of 1536 bp in length
32652	32731: gap of unknown length
32732	34418: contig of 1687 bp in length
34419	34498: gap of unknown length
34499	36084: contig of 1506 bp in length
36085	36270: gap of unknown length
36271	38270: contig of 2186 bp in length
38272	38350: gap of unknown length
38351	40033: contig of 1683 bp in length
40034	40113: gap of unknown length
40114	41483: contig of 1370 bp in length
41484	41563: gap of unknown length
41564	43313: contig of 1750 bp in length
43314	44834: gap of unknown length
43394	44834: contig of 1441 bp in length
44835	44914: gap of unknown length
44915	45960: contig of 1046 bp in length
45961	46040: gap of unknown length
46041	48133: contig of 2093 bp in length
48134	48213: gap of unknown length
48214	49431: contig of 1218 bp in length
49432	49511: gap of unknown length
49512	51678: contig of 2167 bp in length
51679	51758: gap of unknown length
51759	54159: contig of 2401 bp in length
54160	54239: gap of unknown length
54240	56130: contig of 1891 bp in length
56131	56210: gap of unknown length
56211	58314: contig of 2104 bp in length
58315	58394: gap of unknown length
58395	60495: contig of 2101 bp in length
60496	60575: gap of unknown length
60576	62294: contig of 1719 bp in length
62295	62374: gap of unknown length
62375	63520: contig of 1146 bp in length
63521	63500: gap of unknown length
63501	66017: contig of 2417 bp in length
66018	66097: gap of unknown length
66098	67582: contig of 1485 bp in length
67583	67662: gap of unknown length
67663	70566: contig of 2304 bp in length
70567	70546: gap of unknown length
70547	73266: contig of 2620 bp in length
73267	73346: gap of unknown length
73347	75747: contig of 2401 bp in length
75748	75827: gap of unknown length
75828	78728: contig of 2901 bp in length
78729	78808: gap of unknown length
78809	81844: contig of 3036 bp in length
81845	81924: gap of unknown length
81925	86190: contig of 4266 bp in length
86191	86270: gap of unknown length
86271	94031: contig of 7761 bp in length

FEATURES	source
REFERENCE	1 (bases 1 to 1922)
AUTHORS	Seo, J. W., Walter, L. and Gunther, E.
TITLE	Genomic analysis of MIC genes in <i>Rhesus macaques</i>
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1922)
AUTHORS	Walter, L.
TITLE	Direct Submission
JOURNAL	Submitted (18-OCT-2000) Walter L., Division of Immunogenetics, University of Goettingen, Heinrich-Dueker-Weg 12, 37073 Goettingen, GERMANY
FEATURES	Location/Qualifiers
source	1. .1922
mrna	/organism="Macaca mulatta" /db_xref="taxon:9544" join(<1. .256,526. .804,1392. .1670,1773. .1922) /gene="mic3*02" /product="MHC class I-related protein precursor" 1. .1922
gene	/gene="mic3*02" 1. .256
exon	/number=2 join(<1. .256,526. .804,1392. .1670,1773. .>1922) /gene="mic3*02" /codon_start=1 /product="MHC class I-related protein precursor" /protein_id="CAC14176.1" /db_xref="GI:10945038" /translation="ELSLRYNNTVLSRDSGVSDPLAEGLDSOLFVRYDRETRBRAR PGCGMAEAVLGAKTWDDETGTGNGDLRPTLTHICKKGLSLDIONCELYEPO STGSRHFYIDGERFELSINLATOEMPTVAQSSRAOTLAMPKEETMKNTIYHAMQAT CLKLRIYQKSRVAVRRTAPVNVITSEASEGNIYVTCRASGFYPRIALTWODGV SLNDGQOMGILPDONGTYQTWATRIROGEORFACIMHSGNSHSPVSGKVLV FOSQWLDIPYLVGYAAAAVAFAAIIIVIIYLCKRKTSAAEGP"
intron	257. .525 /gene="mic3*02" /number=2 526. .804 /gene="mic3*02" /number=3 805. .1391 /gene="mic3*02" /number=3 1392. .1670 /gene="mic3*02" /number=4 1671. .1772 /gene="mic3*02" /number=4 1773. .1922 /gene="mic3*02" /number=5
exon	412 a 513 c 599 g 398 t
BASE COUNT	412 a 513 c 599 g 398 t
ORIGIN	
Query Match	8.5%; Score 35.2; DB 93; Length 1922;
Best Local Similarity	47.7%; Pred. No. 3.1;
Matches 103; Conservative	0; Mismatches 113; Indels 0; Gaps 0;
OY	191 gcaagggtgctgcacacgctgctactagtaacagagcgtctgctgagcagcaactgatcat 250
Db	1058 GCAGGAGAGGAAGACCTTCAGGGCCAGGGCGTGGCCCTCGCTCCACAGCTGCCATCCC 1117
OY	251 gcataggggacacacgacagcgtctatttgcttaactcatcctctgtagcggtagg 310
Db	1118 AGAGAGTTCCTCCCTGGCCCAAGCCCGGTGAGTCCACCCCAACATCCCTCTCAGCA 1177
OY	311 cagaagctcggttcgtgtatgctgaggggacaggtcccgacacgctgctgagcttg 370
Db	1178 TCAATGTGGGATCCCAAGCTGTAGAGCCACAGTGTCCCAAGGCCCATCTCTGCTAGCCT 1237
OY	371 gacggcaatgggaacctgcatatacccaagggtctga 406

Db	1238	GGAGGAACTGGGCCCCAGGCTGAGGACAGACTTGCA	1273
RESULT	10		
LOCUS	MMU291499		
DEFINITION	MMU291499 1922 bp DNA	PRI	21-Oct-2000
ACCESSION	Macaca mulatta partial mic3*03 gene for MHC class I-related protein		
VERSION	precursor, exons 2-5.		
KEYWORDS	AJ291499		
SOURCE	MHC class I-related protein precursor; mic3*03 gene.		
ORGANISM	rhesus monkey.		
REFERENCE	Macaca mulatta		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;		
JOURNAL	Cercopithecoidea; Macaca.		
REFERENCE	1 (bases 1 to 1922)		
AUTHORS	Seo, J.-W., Walter, L. and Gunther, F.		
TITLE	Genomic analysis of MHC genes in rhesus macaques		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1922)		
AUTHORS	Walter, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-Oct-2000) Walter L., Division of Immunogenetics,		
	University of Goettingen, Helmut-Dueker-Weg 12, 37073 Goettingen,		
	GERMANY		
FEATURES			
SOURCE	1..1922 location/Qualifiers		
	/organism="Macaca mulatta"		
	/db_xref="taxon:9544"		
mRNA	join(<1..256,526..804,1392..1670,1773..>1922)		
	/gene="mic3*03"		
	/product="MHC class I-related protein precursor"		
gene	1..1922		
	/gene="mic3*03"		
	<1..256		
exon	/gene="mic3*03"		
	/number=2		
	join(<1..256,526..804,1392..1670,1773..>1922)		
CDS	/gene="mic3*03"		
	/codon_start=1		
	/product="MHC class I-related protein precursor"		
	/protein_id="CAC14177.1"		
	/db_xref="GI:10945040"		
	/translation="ELSLRYNVTVLSRDSGVSEFLAEGLHDSQLFVRYDRETRRAR		
	POGMAEAVLGAQKTDPTDGTGTLNKGDRITLTHIGKKGLHSLDIONCEYEDG		
	SGSRHRYDGERFLSLNLTQETVAQSSRAQFLANFEKEDTMKTNTHYHAMQAD		
	CLKSLRRYQKSRVAVRTPAPPMVNIITSEASEGNIITVCRAAGFYPRNIALTWQDQ		
	SLNHDGQMGGILPDONGTQYTWATRIROGEORLACVMEHSGNHHVPSPGVILV		
	FQSQMLDIPYLVGVAANAANAIAFIILIVLCCCKKTSAAEP"		
	257..525		
intron	/gene="mic3*03"		
	/number=2		
exon	526..804		
	/gene="mic3*03"		
	/number=3		
intron	805..1391		
	/gene="mic3*03"		
	/number=3		
exon	1392..1670		
	/gene="mic3*03"		
	/number=4		
intron	1671..1772		
	/gene="mic3*03"		
	/number=4		
exon	1773..>1922		
	/gene="mic3*03"		
	/number=5		
BASE COUNT	415 a 508 c 600 g 399 t		
ORIGIN			

Oy	10	atccattgcaatgccttctcgtggcgctcttttgttaacatccccaacaatatgagacc	69
Dn	99799	ATCCATTGCAACATCATTTTGAGTGTCTTTTTTTGGTCCTGAATAACTGTATAGTAGTGG	99740
Oy	70	accaagaatagaagtcttcaccttcggctaactgccgaaggcatcaccttcctaacag	129
Dn	99739	ACGATTACAGCAAACATCTCGCAGGATCTGTGTAAGAAGAAAGAAAATCTCTGCACGCCG	99680
RESULT	13		
LOCUS	AC004133/c		
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC F5A9 from chromosome I, complete sequence.		
ACCESSION	AC004133		
VERSION	AC004133.3		
KEYWORDS	HTG.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II (bases 1 to 105863)		
AUTHORS	Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C., Shinn,P., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,R., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A., Theologis,A. and Ecker,J.R.		
TITLE	Genomic sequence for Arabidopsis thaliana BAC F5A9 from chromosome I		
JOURNAL REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 105863) Ecker,J.R.		
JOURNAL TITLE	Direct Submission		
JOURNAL	Submitted (11-FEB-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
REFERENCE	3 (bases 1 to 105863) Ecker,J.R.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (26-AUG-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
REFERENCE	4 (bases 1 to 105863)		
AUTHORS	Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C., Shinn,P., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,R., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.		
JOURNAL TITLE	Direct Submission		
JOURNAL	Submitted (30-AUG-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA		
REFERENCE	5 (bases 1 to 105863)		
AUTHORS	Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C., Shinn,P., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,R., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.		
JOURNAL TITLE	Direct Submission		
JOURNAL	Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA		
COMMENT	On Aug 26, 2000 this sequence version replaced gi:9625217. Location/Qualifiers		
FEATURES	1..105863		
SOURCE	1..105863		

CDs

/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F5A9"
complement(1003..1779)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F5A9.1"
/protein_id="AAG01131.1"
/db_xref="gi:9945094"
/translation="MSCRGIRPDTORRHRAAROLTTRPKYPARSRVSSSSSAHPSQOAR
OLSPKYPASSSSSSSAHPSQAROLSPKYPASNSVSSSSSAHPSQAROLTTRK
YPASSVSSSSSAHPSQADROLTRKLOKRLSPGASQAROLTRNLGVSP
LRSPAPSGTAROLTTRPKYPASRFVSSSSSAHPSQAROLTTRKOLVSPVSSSS
AHPSQAARLTTRKLYCPSLSSFTSPSSLRAPRK"
join(12177..2246,2688..3020,4764..4927,4977..5060)
/note="unknown protein; similar to ESTs dbj|AVS30941.1,
dbj|AVS50975.1, gb|BE037999.1, gb|BE037993.1,
gb|A1995009.1, gb|A1099531.1, gb|N37906.1, gb|F76894.1"
/codon_start=1
/evidence=not_experimental
/product="F5A9.2"
/protein_id="AAG01107.1"
/db_xref="gi:9945070"
/translation="MAAPVVVAPAINLQHMDLPEGNEDVAAVLTOSTQRMILNAEWSII
LYNNILPESNSFTPIREGLYKVDNRFNQDSNTMRNTGAPSTCKNAIFVDAIT
TGLYFDDGDQMDVAEYOGWVLCRKTRVRHGCGTRLSLYKVPEDROKMKIHMA
MEDMKDEESLDFVSLRQLTTTDTLGTGAFSSVYKRVLEHLRGSIPDA"
join(6532..6629,7268..7338,7427..7464,7543..7653,
7769..7868,7964..8099,8187..8291)
/note="similar to anchorilate synthase beta chain
p1rj02340; similar to ESTs dbj|AVS40153.1,
dbj|AVS57490.1, gb|A197696.1, gb|AW004516.1,
dbj|AVS21371.1"
/codon_start=1
/evidence=not_experimental
/product="F5A9.3"
/protein_id="AAG03108.1"
/db_xref="gi:9945071"
/translation="MAESNSIPSVYNSKONGPIVNDSEFTYNAFCWGLGLO
FEVYNDLITVELKRRKRGVILSPGCGPDGSLIOTVLEGLPIPLFCWGLGLO
CIGAFGCKYRSPGVHGGSSMYHDEKEBEGPSLSMPYVGRHSLVTKDPTF
PSDELVTAMTDEGLVMAARHKRYKHIDGVFHPESITITGKTVIRFKYVKKDS
EKL"

CDs

complement(join(9015..10041,10121..10317,11010..11154,
11221..11285,11380..11481,11571..11777,11965..12273),
12375..12487,12874..12937,13027..13665)
/note="unknown protein; similar to ESTs |AV442495.1,
gb|BE522756.1, gb|T42945.1, gb|BE525268.1, gb|BE522201.1,
gbj|AV526198.1, gb|T42945.1, gb|AA651384.1, dbj|AVS32951.1,
dbj|AVS4151.1, dbj|AV563142.1"
/codon_start=1
/evidence=not_experimental
/product="F5A9.4"
/protein_id="AAG03130.1"
/db_xref="gi:9945093"
/translation="WLRKHGGQVYLARGMEPVYNSKTHDYLDMYTLGVSYONNSY
NLIATGQYHDFQIYVINSNMRIIDATLDFKGYIGRVSLSKGTWYVWASGEERK
RIGKLIIFEDYTERFERLCIPNKYPCGATLAVREKELSMISORIKSAIWTY
NKIGTGVNSMYVALDLPKRCGEGSLYDEKRVVCCONIROGKTDSNIN
PRYHKOKGRRIIVEDYSMRIPVTYAAKRPFLVIRSSAASPTVSLNSGRL
QOTLAGSVYKGLSHSGKSFVYKLNPIAAGRPFERSRIFASIEFADPSLCTTT
LTKDLKIRYEBHLISLEAKGVNDCKRIQIESSDREVEVPIFDGSAKFWADTGG
VGNAQNHDESEYKVMVAHVNYVCNDFVAAAPALERTITGIDFQVAYALOGG
QWFSRNPIDHESFADIASSTRFYVEVEEMRREALIGSLDNAIYVSAHEMMAMP
PLRFDEACRKHILDLIGDLSVSGGNGJPAVHAYKKEEVSSTGEENETMD
STEENEYVSTEAERGDEPHEDSNISESAILQIMERNONGSGKKEEPTNDS
ALGQIDLNNOVEISRYVHCDLLCYAKDNRVWMDYLGQFWIRPNSNGDST
ALGQIINKNKLIRWQIRSVYRTEYIDLRNSMRVLETPGMDONPHLDVGVSV
KGNVTFRAHEHSSGEIDEDDIIIDLEDFLFCFPTTEFLRLPLPHSTDAVTFV
SCVDROGAFLVYHNGLHSDFRTVEVWATTSLEPNSVSKRLTYDMRLATGVRL
FNDVROGAFFLVYHNGLVYFDLQGLSTESARVHTAETISGDKDFKVTYGVAPNNG

[illegible]

RESULT 15
MMU242441

LOCUS	1922 bp	DNA	PRI	24-FEB-2000
DEFINITION	Macaca mulatta partial	MIC3*01 gene, exons 2-5.		
FEATURES				

ACCESSION	AJ242441
VERSION	AJ242441.2
KEYWORDS	MHC class I related protein precursor; MIC3 gene; MIC3*01 allele.
SOURCE	rhesus monkey.
ORGANISM	Macaca mulatta

REFERENCE 1 (bases 1 to 1922)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:02:27 ; Search time 257.37 Seconds
(without alignments)
936.790 Million cell updates/sec

Title: US-09-619-643-2

Perfect score: 413

Sequence: 1 agctgttaacatcattgca.....ccccagggctgaaccgag 413

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.2	7.6	653	21	F14709
2	31.2	7.6	28295	20	X20507
3	30.8	7.5	16812	19	V62175
4	30.4	7.4	1413	21	254316
5	30.4	7.4	1413	21	254317
6	30.4	7.4	34827	21	A81481
7	30.4	7.4	34980	21	F21611
8	30.4	7.4	34980	21	F21612
9	30.4	7.4	837096	21	A81489
10	29.8	7.2	584	21	C01502
11	29.8	7.2	1224	12	Q14369

C	12	29.6	7.2	3215	15	Q70739
C	13	29	7.0	1550	21	F21112
C	14	29	7.0	1550	21	A34990
C	15	29	7.0	3089	21	F21113
C	16	29	7.0	3089	21	A34991
C	17	28.6	6.9	2160	20	223134
C	18	28.6	6.9	2509	21	C78146
C	19	28.4	6.9	1176	8	N70133
C	20	28.4	6.9	1443	16	Q94110
C	21	28.4	6.9	1486	16	T03942
C	22	28.4	6.9	1486	16	T04049
C	23	28.4	6.9	1486	16	Q97690
C	24	28.4	6.9	1486	17	T34851
C	25	28.4	6.9	1486	17	T37382
C	26	28.4	6.9	1486	17	T32590
C	27	28.4	6.9	1486	21	A51992
C	28	28.4	6.9	1486	21	237784
C	29	28.4	6.9	1536	16	Q94111
C	30	28.4	6.9	2169	15	O56615
C	31	28.2	6.8	434	21	C52253
C	32	28.2	6.8	1059	21	Z38553
C	33	28.2	6.8	1317	15	O66179
C	34	28.2	6.8	1317	19	V18357
C	35	28.2	6.8	1317	21	A91726
C	36	28.2	6.8	1356	21	A07598
C	37	28.2	6.8	1588	20	Z27610
C	38	28.2	6.8	1664	21	A61656
C	39	28.2	6.8	1664	21	Z40014
C	40	28.2	6.8	1737	13	Q29506
C	41	28.2	6.8	1737	16	O80521
C	42	28.2	6.8	1737	16	O99007
C	43	28.2	6.8	1944	20	X15882
C	44	28.2	6.8	2084	21	C77109
C	45	28.2	6.8	5161	21	F20896

ALIGNMENTS

RESULT 1
F14709/c
ID F14709 standard; cDNA; 653 BP.
XX
AC F14709;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO: 7232.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PR substrate of expressed sequence tags -
 PS Claim 88; Page 2930; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway engineering.
 CC Using ESTs provides several advantages over genomic or random cDNA
 CC clones including elimination of redundancy as one spot on an array
 CC equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. P07478 to F11247 represents ESTs from *Fusarium*
 CC *venenatum*; F11248 to F11853 represents ESTs from *Aspergillus niger*;
 CC F11854 to F14878 represents ESTs from *Aspergillus oryzae*; and F14879 to
 CC F15337 represents ESTs from *Trichoderma reesei*, which are all
 CC specifically claimed in the present invention.

SO Sequence 653 BP; 134 A; 134 C; 173 G; 210 T; 2 other;

Query Match 7.6%; Score 31.2; DB 21; Length 653;
 Best Local Similarity 46.4%; Pred. No. 0.99;
 Matches 102; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

OY 48 accctccacaatatatgaccaccatgaatgaagatcccttcgctatgctccga 107
 DB 483 AACCTCGCAATGAAAGATTCGAATTTATAGATTAACCTTGATCCAGATTCCTCCCG 423
 OY 108 aaggctaatcccttcataacagcgctgaagctcagcgatctactacatgatatcaat 167
 DB 422 CAGACAAAACCAACAGATCAAAAATACGACACATCCAAATCAATGCTTTGAAG 363
 OY 168 aggcacctctgcgacacatgagcgaggtgagctgcacgctgcttagtacagagcgt 227
 DB 362 AACCAATTTGACTATACCTCCACAGTCTGCGCGCAGCAGAGAGATTTCACA 303
 OY 228 ctggctggccagcaactgctatgcatagtaggacacatgac 267
 DB 302 CTTGCTGCACGTTATGCGCTTAATTCGCCAGGCCCTCCGC 263

RESULT 2

X20507/c
 ID X20507 standard; DNA; 28295 BP.

AC X20507;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of *Treponema pallidum*.

KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;

KW enzyme production; ds.

OS *Treponema pallidum*.

PN W09859034-A2.

PD 30-DEC-1998.

XX

PF 23-JUN-1998; 98WO-US13041.

XX 24-JUN-1997; 97US-0050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

DR WPI; 1999-081273/07.

PT New isolated *Treponema pallidum* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. *pallidum* infections, particularly syphilis

PS Claim 1; Page 210-227; 1150pp; English.

CC X20500-21243 represent polynucleotide sequences from the genome of
 CC *Treponema pallidum*. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. *pallidum*
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to *Borrelia* infections in animals, and for the
 CC production of biosynthetic products such as enzymes.

SO Sequence 28295 BP; 6269 A; 6077 C; 8919 G; 6996 T; 34 other;

Query Match 7.6%; Score 31.2; DB 20; Length 28295;
 Best Local Similarity 55.0%; Pred. No. 6.2;
 Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 301 gccggtagcgcagagagctcggtgctgagctcgagggcagagttccgcagcgtgtgc 360

DB 5230 GCCCGTATCATGAGGCGACAGGTCTCAGCAGGACGCGAACCGCTTCGCGCGGT 5171

OY 361 ttgacgttcggacggcaatggagctgcataccccaaggtcggaacc 409

DB 5170 CTTGCTGCGCANACAAACGTTATACGGGCCAGGCCACAGCGCACGAAAC 5122

RESULT 3

V62175
 ID V62175 standard; DNA; 16812 BP.

AC V62175;

DT 08-JAN-1999 (first entry)

DE HSV-2 strain SBS Contlg ID 12 DNA sequence.

KW HSV-2 strain SBS; Immunological response induction; therapy;

KW antiviral identification; viral protein inhibitor; ss.

OS Herpes simplex virus type 2.

PH location/Qualifiers

FT 127..1371

FT /tag= a

FT /product= "ORF#1 protein"

FT /note= "encoded protein shown in W72159"

FT complement (1553..2428)

FT /tag= b

FT /product= "ORF#2 protein"

FT /note= "encoded protein shown in W72160"

FT 2714..4159

FT /tag= c

FT /product= "ORF#3 protein"

FT /note= "encoded protein shown in W72161"

FT 6835..6948

FT /tag= d

FT /product= "ORF#4 protein"

FT /note= "encoded protein shown in W72162"

FT 7392..8573

FT /tag= e

FT CDS

FT	/product= "ORF#5 protein"
FT	/note= "encoded protein shown in W72163"
FT	8775..9893
FT	/*tag= f
FT	/product= "ORF#6 protein"
FT	/note= "encoded protein shown in W72164"
FT	10212..11858
FT	/*tag= g
FT	/product= "ORF#7 protein"
FT	/note= "encoded protein shown in W72165"
FT	12010..12147
FT	/*tag= h
FT	/product= "ORF#8 protein"
FT	/note= "encoded protein shown in W72166"
FT	12247..12516
FT	/*tag= i
FT	/product= "ORF#9 protein"
FT	/note= "encoded protein shown in W72167"
FT	complement (13004..13912)
FT	/*tag= j
FT	/product= "ORF#10 protein"
FT	/note= "encoded protein shown in W72168"
FT	15899..16582
FT	/*tag= k
FT	/product= "ORF#11 protein"
FT	/note= "encoded protein shown in W72169"
XX	
PX	W09820016-A1.
PD	14-MAY-1998.
XX	
PE	31-OCT-1997; 97MO-US20016.
XX	
PR	09-JUN-1997; 97US-0049018.
PR	04-NOV-1996; 96US-0030279.
XX	
PA	(SMIK) SMITHKLINE BEECHAM CORP.
PI	Chan JY, Debrowski-Amara CE, Delvecchio AM, Dillon SB;
PI	Esser KM, Leary JJ;
XX	
DR	WP1: 1998-286847/25.
DR	P-SDB: W72159, W72160, W72161, W72162, W72163, W72164, W72165, W72166,
DR	W72167, W72168, W72169.
XX	
PT	Herpes simplex virus type-2 sequences - useful in, e.g., prevention
PT	and treatment of infection or inducing immunological response in
PT	mammal
PS	Claim 1; Page 505-512; 748pp; English.
XX	
CC	This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC	sequence of the invention. This sequence was isolated from HSV-2 strain
CC	SBS (deposited as ATCC VR-2546), is designated Config ID 12, and encodes
CC	11 HSV-2 proteins. The proteins can be used for the treatment or
CC	prevention of disease, to induce an immunological response in a mammal or
CC	to identify inhibitors, activators or novel antivirals. Antagonists of
CC	the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC	or a vector containing it can also be used to induce an immunological
CC	response in a mammal.
XX	
SO	Sequence 16812 BP; 2708 A; 5989 C; 5367 G; 2748 T; 0 other;
XX	
Query Match	7.5%; Score 30.8; DB 19; Length 16812;
Best Local Similarity	46.7%; Pred. No. 6.6; Mismatches 112; Indels 0; Gaps 0
Matches 98; Conservative	0;
OY	193 aggggtgagctgcacacgttcctagtacaaggacgtcgtgctgcccagcacatcatcgtc 252
	I 11
16256 atggttgagccagcgacctgcgaacatccgtcgtcccccccgagagcgagcgcgcgaa 16315	
OY	253 ataggggcaacactgcagagccttatgtgtaacttaaacatccatcccttgaagccggtatgcga 312

Db 16316 gaagaagcagaccgcctcgcttctccgcccggacccccctcatgtgtgccgatgaaggc 1637
||| | ||| |||| | | | | | | | | |
Oy 313 gaggtcgcgttgtagctgtgagcgtgagggcacagatctctgcacggtgtcgttaacgtcoga 372
||| | | | | | | | | | | | | | | | | |
Db 16376 gaggaagaacagcacgacgacgatgaacgaacacgcgacgcggcgcttggtccgcgga 1643
| | | | | | | | | | | | | | | | | | | |
Oy 373 cgcgaatggagacctgtgcattcccccaaggc 402
| | | | | | | | | | | | | | | | | | | |
Db 16436 ccggagaacagactccgcggtccgcggggc 16465
| | | | | | | | | | | | | | | | | | | |

RESULT 4

ID Z54316/C DNA; 1413 BP.

XX Z54316;

AC 254316;

XX DT 21-MAR-2000 (first entry)

XX NEisseria meningitidis ORF 752 partial DNA sequence SEQ ID NO:2581.

DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen: Vaccine:

KW Antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

KM antibacterial; gene therapy; ds.

OS Neisseria meningitidis.

PN W09957280-A2.

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzato M, Rappunli R, Ratli G, Scalato E, Scarselli M,
PI Tettelin H, Venter JC;
DR WPJ: 2000-062150/05.
DR P-PSDB; Y75554.

PT Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
PS Claim 7; Page 1227; 1453pp; English.

CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrhoeae polynucleotides and
CC polypeptides. Z54537 to Z54576 and Z54616 to Z5473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenetic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC have been used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

SO Sequence 1413 BP; 445 A; 315 C; 304 G; 349 T; 0 other;

Query Match	7.4%	Score 30.4	DB 21	Length 1413
Best Local Similarity	54.5%	Pred. No. 2.7		
Matches 61	Conservative 0	Mismatches 51	Indels 0	Gaps 0
Oy	68	ccacatgaatgaagagatccctccctccgtatgctccgcgaagagctaaccttcaatc	127	
Db	781	CCATCAGCGTATGAAACCTCTCCGTCGGCGGGTGGTTGATACAGGCGCTTACCATTCATATAT	722	
Oy	128	agggcgctgaagctccagcagctactacaatgatalatacaatagcaccctgtc	179	
Db	721	CGCGCATTAAGATTTGCGTATCTCCTCGTGAATTTGCCGGGCTCGGCTTGT	670	
RESULT 5				
Z54317/c				
ID	Z54317	standard	DNA	1413 BP.
XX				
XX	Z54317/			
XX				
DT	21-MAR-2000	(first entry)		
XX				
DE	Neisseria meningitidis ORF 752 partial DNA sequence SEQ ID NO:2583.			
XX				
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;			
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;			
KW	antibacterial; gene therapy; ds.			
XX				
OS	Neisseria meningitidis.			
XX				
PN	W09957280-A2.			
PD				
XX	11-NOV-1999.			
PF	30-APR-1999;	99WO-US09346.		
XX				
PR	01-MAY-1998;	98US-0083758.		
PR	31-JUL-1998;	98US-0094869.		
PR	02-SEP-1998;	98US-0098994.		
PR	02-SEP-1998;	98US-0099062.		
PR	09-OCT-1998;	98US-0103749.		
PR	09-OCT-1998;	98US-0103794.		
PR	09-OCT-1998;	98US-0103796.		
PR	25-FEB-1999;	99US-0121528.		
XX				
PA	(CHIR) CHIRON CORP.			
PA	(GENO-) INST GENOMIC RES.			
XX				
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;			
PI	Peterelin J, Pizzza M, Rappuoli R, Ratel G, Scalato E, Scarselli M;			
PI	Tetzelin H, Venter JC;			
XX				
DR	WPI: 2000-062150/05.			
DR	P-PSDB: Y75555.			
XX				
PT	Novel Neisserial polypeptides predicted to be useful antigens for			
PT	vaccines and diagnostics			
XX				
PS	Claim 7; Page 1227-1228; 1453bp; English..			
XX				
CC	Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent			
CC	novel Neisseria meningitidis and N. gonorrhoeae polynucleotides and			
CC	polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR			
CC	primers used in the exemplification of the present invention. The			
CC	polypeptides, the polynucleotides, antibodies and compositions of			
CC	the invention can be used as vaccines, as diagnostic reagents, and as			
CC	immunogenic compositions. The polypeptides can be used in the			
CC	manufacture of medicaments for treating or preventing infection due to			
CC	Neisserial bacteria (e.g. meningitis and septicaemia), to detect the			
CC	presence of Neisseria bacteria, or to raise antibodies. They may also			
CC	be used to screen for agonists or antagonists, which may themselves			
CC	have use as antibacterial agents. The polynucleotides of the invention			

CC may also be used in gene therapy protocols.

XX

SQ Sequence 1413 BP; 445 A; 315 C; 304 G; 349 T; 0 other;

Query Match 7.4%; Score 30.4; DB 21; Length 1413;
Best Local Similarity 54.5%; Pred. No. 2.7;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0

Oy 68 ccacatgaataagaagctccctcggcatgctcccgaaaggccaatccctcaaac 127
||| || | |||| | ||| |||| | | |||| | ||| |||
Db 781 CCATCAGCGCTATCAACCTGTCCTGGGGCGGTGTGATACAGGCTTTCACATTGATAT 722

Oy 128 aggcgtcgaagctccagactacaatatatatacaaataggcaccttgc 179
|||| | ||| | | ||| | ||| | |||| |
Db 721 CGGCATATAAAGATTTCGTATCCTGCTGATATGTCGGGCTCGCCTGT 670

RESULT 6
AB1481
ID AB1481 standard; DNA; 34827 BP.
XX
AC AB1481;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_29 SEQ ID NO:29.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
antigen; vaccine; diagnosis; infection; antibacterial; identification;
Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
PN WO200022430-A2.
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
PI Frazer CM, Hickey E, Peterson J, Pettelin H, Venter JC:
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcelli M, Scariato V;
PI Rappelli R, Pizzo M;
XX
DR WPI: 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 572-582; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AB1453 to AB2414 represent
CC specifically claimed Neisseria meningitidis genomic DNA sequences. AB1260
CC to AB1303 and B25620 to B25663 represent Neisseria DNA sequences and
CC their corresponding proteins: AB1254 to AB1259 and AB1304 to AB1321
CC represent PCR primers used in the isolation of Neisseria meningitidis DNA
CC sequences; and AB1322 to AB1452 represent Neisseria meningitidis MenB
CC polynucleotide ORF sequences, which are all used in the exemplification
CC of the present invention. The nucleic acid sequences, protein sequences,
CC composition. The composition can be used as a medicament (or in the
CC manufacture of a medicament) for treating, preventing or diagnosing
CC infection due to Neisserial bacteria. For example, some of the identified
CC proteins could be components of vaccines against Meningococcus B; against
CC all serotypes; and/or against all pathogenic Neissariae. Identification
CC of sequences from the bacterium will also facilitate production of

The present invention describes the full length genome of *Neisseria meningitidis* B (NMb). The sequences in F21544 and F21607 to F21613 represent fragments of the NMb genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of F21544 is repeated at the beginning of F21607, the last 49980 bp of F21607 are repeated at the beginning of F21608, and so on). F21545 to F21588 encode the *Neisseria* proteins given

PT Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
CC *Neisseria meningitidis* B (NMB). The sequences in F21544 and F21607 to
CC F21613 represent fragments of the NMB genomic sequence, as the sequence

CC was too long to go in a record on its own it was split into 8 sequences
CC which overlap each other at the beginning and end of each sequence by
CC 49980 bp (i.e. the last 49980 bp of F21544 is repeated at the beginning
CC of F21607, the last 49980 bp of F21607 are repeated at the beginning of
CC F21608, and so on). F21545 to F21588 encode the *Neisseria* proteins given
CC in B58550 to B58593, and F21589 to F21606 represent PCR primers which
CC are used in the exemplification of the present invention. The MNB genome
CC and fragments from it have antibacterial activity, and can be used in
CC vaccines and gene therapy. *Neisseria* nucleic acids, proteins and/or
CC antibodies which bind to the proteins can be used in compositions for
CC treating or preventing infection due to *Neisseria* bacteria or as a
CC diagnostic reagent for detecting the presence of *Neisseria* bacteria or
CC of antibodies raised to *Neisseria* bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the MNB
CC genome. The DNA sequences provide further opportunities to find
CC antigenic or immunogenic proteins which are more effective in vaccines
CC than the outer membrane proteins currently used.

SO Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Query Match 7.4%; Score 30.4; DB 21; Length 349980;

Best Local Similarity 54.5%; Pred. No. 39;

Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 68 ccaccctgaatgaagatcctcctcgctatgctccgaaagcctaaccctcaatc 127

DB 38135 CCATCAGCGGTATGACCTGCGTGGCGGTGTGATACAGCGCTTACCATGTAT 38076

OY 128 aggcgtgaagctccagcgatactacaatgataatagcaccctgct 179

DB 38075 CGCGCATTAAGATTTCGTATCTCCTCGCAATTGTCGGGCTCGGCTTGTT 38024

RESULT 9
A81489/c
ID A81489 standard; DNA: 837096 BP.

AC A81489;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.

KM *Neisseria meningitidis*; *Neisseria gonorrhoeae*: genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW *Meningococcus B*; MemB; ds.

OS *Neisseria meningitidis*.

PN W0200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratli G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;

DR WPI: 2000-318079/27.

XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
XX used in the diagnosis and treatment of *N. meningitidis* infection and
XX other *Neisseria* infections, for example, *N. gonorrhoea* -

PS Claim 7; Page 629-865; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
CC proteins from *Neisseria* genomic sequences. A81453 to A82414 represent
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences; A81260
CC to A81303 and B25620 to B25663 represent *Neisseria* DNA sequences and
CC their corresponding proteins; A81254 to A81259 and A81304 to A81321
CC represent PCR primers used in the isolation of *Neisseria meningitidis* DNA
CC sequences; and A81322 to A81452 represent *Neisseria meningitidis* MNB
CC polynucleotide ORF sequences, which are all used in the exemplification
CC of the present invention. The nucleic acid sequences, protein sequences,
CC and antibodies against them, can be used in the manufacture of a
CC composition. The composition can be used as a medicament (or in the
CC manufacture of a medicament) for treating, preventing or diagnosing
CC infection due to *Neisseria* bacteria. For example, some of the identified
CC proteins could be components of vaccines against *Meningococcus B*; against
CC all serotypes; and/or against all pathogenic *Neisseriae*. Identification
CC of sequences from the bacterium will also facilitate production of
CC biological probes, particularly organism-specific probes. Attempts to
CC make efficacious *Meningococcus B* vaccines have failed mainly due to
CC antigen tolerance. Multivalent vaccines have also been tried but none
CC have successfully overcome antigenic variability. The provision of
CC further, complete sequences may provide an opportunity to identify
CC secreted or surface exposed proteins that may be presumed targets for the
CC immune system and which are not antigenically variable or at least more
CC conserved than other more variable regions.

SO Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

Query Match 7.4%; Score 30.4; DB 21; Length 837096;

Best Local Similarity 54.5%; Pred. No. 57;

Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 68 ccaccctgaatgaagatcctcctcgctatgctccgaaagcctaaccctcaatc 127

DB 402906 CCATCAGCGGTATGACCTGCGTGGCGGTGTGATACAGCGCTTACCATGTAT 402847

OY 128 aggcgtgaagctccagcgatactacaatgataatagcaccctgct 179

DB 402846 CGCGCATTAAGATTTCGTATCTCCTCGCAATTGTCGGGCTCGGCTTGTT 402795

RESULT 10

ID C01502 standard; CDNA: 584 BP.

AC C01502;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 1500.

KW Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; G01496.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

CC lactobacillus. The protein is used in the prepn. of fermented

PT cancers -
XX
PS Disclosure: Page 888; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. A32313 to A35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680 (A32323
CC to A33992) are specifically claimed ONs from the present invention.
CC N.B. Sequences given in the disclosure of the present invention do not
CC match up with their corresponding SEQ ID NO: sequences given in the
CC sequence listing.
XX
SQ Sequence 1550 BP; 292 A; 454 C; 549 G; 255 T; 0 other;
XX
Query Match 7.0%; Score 29; DB 21; Length 1550;
Best Local Similarity 55.4%; Pred. NO. 8.5;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
XX
QY 311 cagaagtcggttgtagctgagctggagggcagagttccctgcacccggtgtagctgacgtcg 370
DB 153 ccgcgggcccgtgtgtcccccgggggaggggagggcgtgcgacgacctggtcggggtcgg 212
QY 371 gacggcaatggagacctggacatacccccagggcgtggaaccgg 411
DB 213 gccgggggtcggacactggagacatccctcgtgaggaagggccgg 253
XX
RESULT 15
F21113
ID F21113 standard; DNA; 3089 BP.
XX
AC F21113;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2680.
XX
DE Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX human; airway disorder; bronchoconstriction; lung inflammation;
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
XX respiratory obstruction; pulmonary obstruction; Impeded respiration;
XX surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200062736-A2.
XX
XX 26-OCT-2000.
XX
XX

PF 24-MAR-2000; 2000WO-US08020.
XX
XX 06-APR-1999; 99US-0127958.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX (NYCE/) NYCE J W.
XX
XX NYCE JW;
XX
XX WPI; 2000-679539/66.
XX
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX
XX Disclosure: Page 963-964; 1592pp; English.
XX
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'universal' or alternative base.
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasoactive peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
XX condition selected from pulmonary vasoconstriction, inflammation,
XX allergies, asthma, Impeded respiration, respiratory distress syndrome
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,
XX and/or cancer. F18434 to F21543 represent human polynucleotide fragments
XX and antisense oligonucleotides used in the exemplification of the
XX present invention.
XX
XX
XX Sequence 3089 BP; 650 A; 841 C; 971 G; 627 T; 0 other;
XX
XX
Query Match 7.0%; Score 29; DB 21; Length 3089;
Best Local Similarity 55.4%; Pred. NO. 12;
Matches 56; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
XX
QY 311 cagaagtcggttgtagctgagctggagggcagagttccctgcacccggtgtagctgacgtcg 370
DB 1692 ccgcgggcccgtgtgtcccccgggggaggggagggcgtgcgacgacctggtcggggtcgg 1751
QY 371 gacggcaatggagacctggacatacccccagggcgtggaaccgg 411
DB 1752 gccgggggtcggacactggagacatccctcgtgaggaagggccgg 1792
XX
Search completed: May 8, 2001, 15:06:20
Job time: 42921 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:01 ; Search time 132.33 Seconds

(without alignments)
544,932 Million cell updates/sec

Title: US-09-619-643-2

Perfect score: 413
Sequence: 1 agctcgttaacatcattgca.....ccccaggcgtgaaccgag 413

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28.4	6.9	1486	1 US-08-250-859-1	Sequence 1, Appl
C 2	28.4	6.9	1486	1 US-08-347-029-1	Sequence 1, Appl
C 3	28.4	6.9	1486	1 US-08-490-803-1	Sequence 1, Appl
C 4	28.4	6.9	1486	2 US-08-457-254-1	Sequence 1, Appl
C 5	28.4	6.9	1486	2 US-08-484-257-1	Sequence 1, Appl
C 6	28.4	6.9	1486	3 US-08-999-927-3	Sequence 3, Appl
C 7	28.4	6.9	1486	5 PCT-US94-08806-1	Sequence 1, Appl
C 8	28.4	6.9	1486	5 PCT-US95-01775-1	Sequence 1, Appl
C 9	28.4	6.9	1486	5 PCT-US95-01829-3	Sequence 3, Appl
C 10	28.4	6.9	1486	5 PCT-US95-14932-1	Sequence 1, Appl
C 11	28.4	6.9	1486	5 PCT-US95-16626-1	Sequence 1, Appl
C 12	28.4	6.9	2169	1 US-08-379-496-1	Sequence 4, Appl
C 13	28.4	6.8	1317	1 US-08-153-848-45	Sequence 45, Appl
C 14	28.2	6.8	1317	5 PCT-US93-11153-45	Sequence 45, Appl
C 15	28.2	6.8	1317	5 PCT-US93-11153-45	Sequence 45, Appl
C 16	28.2	6.8	1737	1 US-08-202-056-4	Sequence 4, Appl
C 17	28.2	6.8	1737	1 US-08-076-093A-3	Sequence 3, Appl
C 18	28.2	6.8	1737	1 US-08-701-265-3	Sequence 3, Appl
C 19	28.2	6.8	1737	2 US-08-284-586-3	Sequence 3, Appl
C 20	28.2	6.8	1737	2 US-08-805-478-3	Sequence 3, Appl
C 21	28.2	6.8	1737	2 US-08-802-627A-3	Sequence 3, Appl
C 22	28.2	6.8	1737	2 US-08-801-238-3	Sequence 3, Appl
C 23	28.2	6.8	1737	2 US-08-801-228-3	Sequence 3, Appl
C 24	28.2	6.8	1737	3 US-09-104-296-3	Sequence 3, Appl
C 25	28.2	6.8	1737	5 PCT-US94-06380-2	Sequence 5, Appl
C 26	28.2	6.8	6714	1 US-08-021-623C-5	Sequence 42, Appl
C 27	27.6	6.7	1256	4 US-09-318-448-42	Sequence 42, Appl

28	27.6	6.7	1273	4 US-09-318-448-45	Sequence 45, Appl
29	27.6	6.7	1275	4 US-09-318-448-41	Sequence 41, Appl
30	27.6	6.7	1908	4 US-09-318-448-36	Sequence 36, Appl
31	27.4	6.6	4039	1 US-08-363-300-1	Sequence 1, Appl
32	27.2	6.6	5420	6 5256642-3	Patent No. 5256642
33	27.2	6.6	5420	6 5472939-3	Patent No. 5472939
34	27.2	6.6	6951	6 5256642-1	Patent No. 5256642
35	27.2	6.6	6951	6 5472939-1	Patent No. 5472939
C 36	26.8	6.5	1337	3 US-08-946-026-17	Sequence 17, Appl
C 37	26.8	6.5	1797	3 US-08-946-026-13	Sequence 13, Appl
C 38	26.6	6.4	1080	1 US-08-137-627-3	Sequence 3, Appl
C 39	26.6	6.4	1080	2 US-08-865-348-3	Sequence 3, Appl
C 40	26.6	6.4	1641	2 US-08-792-824-5	Sequence 5, Appl
C 41	26.6	6.4	3360	1 US-07-596-467-3	Sequence 3, Appl
C 42	26.6	6.4	3360	1 US-07-934-374-3	Sequence 3, Appl
C 43	26.6	6.4	3360	1 US-07-783-861C-1	Sequence 1, Appl
C 44	26.6	6.4	4771	4 US-08-840-062-3	Sequence 3, Appl
C 45	26.4	6.4	2168	3 US-08-749-522-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-250-859-1/c
Sequence 1, Application US/08250859
Patent No. 5541085
GENERAL INFORMATION:
APPLICANT: Holly, Richard D.
TITLE OF INVENTION: METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/250,859
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E 31-648
REGISTRATION NUMBER: 94-2
REFERENCE/DOCKET NUMBER: 94-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-632-4009
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: 1081
FEATURE:
NAME/KEY: CDS
LOCATION: 105..1241
US-08-250-859-1
Query Match 6.9%; Score 28.4; DB 1; Length 1486;
Best Local Similarity 49.3%; Pred. No. 4.5;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY	187	tggagcaaggtgaggtctggcaacgctggctatgaaggacgctgctgagccagcaatga	246
Db	1330	tggagcagatgacacgtccctctcagcccttctggggaaaccttgctcccgaggaagctgc	1271
QY	247	tcaatgcaatgaggcaactgcagcctcttatgtgcaatgaactcactcctgtgaagcgtg	306
Db	1270	agacgcctcacttgggccagttcccccgcctatggttccctgaacaaattcctggagatgaagg	1211
QY	307	agcgcaaggtccggtctcgatgaactcggaag	336
Db	1210	tacattttgactggatgagggcgctgagag	1181

```

US-08-347-029-1/c
: Sequence 1, Application US/08347029
: Patent No. 5641655
:
: GENERAL INFORMATION:
:   APPLICANT: Foster, Donald C.
:   APPLICANT: Heipel, Mark
:   APPLICANT: Holly, Richard D.
:   TITLE OF INVENTION: METHOD FOR SECRETING THROMBOPOIETIN
:   TITLE OF INVENTION: POLYPEPTIDES
:   NUMBER OF SEQUENCES: 15
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Zymogenetics, Inc.
:     STREET: 1201 Eastlake Avenue East
:     CITY: Seattle
:     STATE: WA
:     COUNTRY: USA
:   ZIP: 98102
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent In Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/347, 029
:     FILING DATE:
:     CLASSIFICATION: 435
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Parker, Gary E
:       REGISTRATION NUMBER: 31-648
:       REFERENCE/DOCKET NUMBER: 94-13
:       TELECOMMUNICATION INFORMATION:
:         TELEPHONE: 206-442-6673
:         TELEFAX: 206-442-6678
:         INFORMATION FOR SEQ ID NO: 1:
:           SEQUENCE CHARACTERISTICS:
:             LENGTH: 1486 base pairs
:             TYPE: nucleic acid
:             STRANDEDNESS: double
:             TOPOLOGY: linear
:             MOLECULE TYPE: cDNA
:             FEATURE:
:               NAME/KEY: CDS
:               LOCATION: 105..1241
:
: US-08-347-029-1

```

[illegible]

```

QY      307  acgcgagaggtccggttcggtagactgag      336
          |   |   |   |   |   |   |   |
Db      1210 TACATTGTGACTGATGAGGGCGGTAGAG      1181

```

```

1      RESULT      3
2      US-08-490-803-1/c
3      : Sequence 1, Application US/08490803
4      : Patent No. 5705349
5      : GENERAL INFORMATION:
6      : APPLICANT: Holly, Richard D.
7      : APPLICANT: Burkhead, Steven K.
8      : TITLE OF INVENTION: METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
9      : NUMBER OF SEQUENCES: 17
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: ZymoGenetics, Inc.
12     : STREET: 4225 Roosevelt Way, N.E.
13     : CITY: Seattle
14     : STATE: WA
15     : COUNTRY: USA
16     : ZIP: 98105
17
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: PatentIn Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/490,803
25
26     : FILING DATE:
27     : CLASSIFICATION: 435
28     : ATTORNEY/AGENT INFORMATION:
29     : NAME: Parker, Gary E
30     : REGISTRATION NUMBER: 31-648
31     : REFERENCE/DOCKET NUMBER: 94-2
32     : TELECOMMUNICATION INFORMATION:
33     : TELEPHONE: 206-547-8080 ext 322
34     : TELEFAX: 206-632-4009
35     : INFORMATION FOR SEQ ID NO: 1:
36     : SEQUENCE CHARACTERISTICS:
37     : LENGTH: 1486 base pairs
38     : TYPE: nucleic acid
39     : STRANDEDNESS: double
40     : TOPOLOGY: linear
41     : MOLECULE TYPE: cDNA
42     : IMMEDIATE SOURCE:
43     : CLONE: 1081
44     : FEATURE:
45     : NAME/KEY: CDS
46     : LOCATION: 105..1241
47
48     US-08-490-803-1

```

Query Match	6.9%	Score 28.4	DB 1	Length 1486
Best Local Similarity	49.3%	Pred. No. 4.5		
Matches	74	Conservative	0	Mismatches 76; Indels 0; Gaps 0.

OY	187	tggcgcaaggttggctctgccaccgttgcttaagacgctcttgcttgccagcaactga	246
Db	1330	tgGAGCAGATGCAGCTGCCTTCACGCTTCCTCGGGGAGCTTGCTCCCGCAGAGAGCTTC	1271
OY	247	tcatcatagaggagcaacacgtcagcgcttatgtgcttaataactcatcctgttaagcgagt	306
Db	1270	AGACGCTCACTGGCCAGTGGCCCCGCGCTATGTCTTCTTGAGACAAATTCTCGGAGATGAGGG	1211
OY	307	agcgcagaggttcggttcggttgagcttgag	336
Db	1210	TACATTGTGACTGATGTAGGGCGGTAGAG	1181

RESULT	4
US-08-457-254-1/C	
: Sequence 1, Application US/08457254	
: Patent No. 5986049	

```

: GENERAL INFORMATION:
: APPLICANT: Forstrom, John W
: APPLICANT: Lofton-Day, Catherine E
: APPLICANT: Lok, SI
: TITLE OF INVENTION: PURIFIED THROMBOPOIETIN AND METHOD OF
: NUMBER OF INVENTION: MAKING IT
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,254
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31,648
: REFERENCE/DOCKET NUMBER: 94-11C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6673
: TELEFAX: 206-442-6678
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 105..1241
: US-08-457-254-1

Query Match
Best Local Similarity 49.3%; Pred. No. 4.5; DB 2; Length 1486;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 187 tggcgaggggtggtgctgcccacgctgctagctagacagcgtctgctgcccagcaactga 246
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1330 TGGAGCAGATGCACCTGCTCTCAGCCTTCTGCGGAGAGCTTGTCCCGAGAGAGCTGC 1271

QY 247 tcatgcataggggacacactgagcgccttatgtgcatcaataacatccatctgtaggcggt 306
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1270 AGAGCCTCAGCTGGGCGCACTGCGCCGCTATGTTCTTCTGAGACAATTCCTGGATGAGGG 1211

QY 307 agcgacagaggtcggttcgctgtagctggag 336
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1210 TACATTGTGACTGATGAGGGCGGCTAGAG 1181

RESULT 5
US-08-484-257-1/c
: Sequence 1, Application US/08484257
: Patent No. 5989537
: GENERAL INFORMATION:
: APPLICANT: HOLLY, Richard D.
: APPLICANT: Lok, SI
: APPLICANT: Foster, Donald C.
: APPLICANT: Hagen, Frederick S.
: APPLICANT: Kaushansky, Kenneth
: APPLICANT: Kujper, Joseph L.
: APPLICANT: Lofton-Day, Catherine E.
: APPLICANT: Oort, Pieter J.
```

```

: TITLE OF INVENTION: Methods for stimulating Granulocyte/Macrophage Lineage Cell
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake E.
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,257
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, Gary E
: REGISTRATION NUMBER: 31-648
: REFERENCE/DOCKET NUMBER: 95-09
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-442-6678
: TELEFAX: 206-442-6678
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1486 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: IMMEDIATE SOURCE:
: CLONE: 1081
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 105..1241
: US-08-484-257-1

Query Match
Best Local Similarity 49.3%; Pred. No. 4.5; DB 2; Length 1486;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 187 tggcgaggggtggtgctgcccacgctgctagctagacagcgtctgctgcccagcaactga 246
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1330 TGGAGCAGATGCACCTGCTCTCAGCCTTCTGCGGAGAGCTTGTCCCGAGAGAGCTGC 1271

QY 247 tcatgcataggggacacactgagcgccttatgtgcatcaataacatccatctgtaggcggt 306
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1270 AGAGCCTCAGCTGGGCGCACTGCGCCGCTATGTTCTTCTGAGACAATTCCTGGATGAGGG 1211

QY 307 agcgacagaggtcggttcgctgtagctggag 336
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1210 TACATTGTGACTGATGAGGGCGGCTAGAG 1181

RESULT 6
US-08-999-927-3/c
: Sequence 3, Application US/08999927
: Patent No. 6099830
: GENERAL INFORMATION:
: APPLICANT: Kaushansky, Kenneth
: TITLE OF INVENTION: Methods of Stimulating Erythropoiesis
: NUMBER OF INVENTION: Using Hematopoietic Proteins.
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/999,927
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/347,748
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 94-09C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: 1081
FEATURE:
NAME/KEY: CDS
LOCATION: 105..1241
US-08-999-927-3
```

```
Query Match      6.9%; Score 28.4; DB 3; Length 1486;
Best Local Similarity 49.3%; Pred. No. 4.5;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 187 tggcgagaggtggtgcacacgtgcttagtacagagcgtctgctgagcagcaactga 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1330 tggacagatgacagcgtcttcacagccttcctggggagaccttgctcccgagaaactgc 1271

QY 247 tcatgcataggggacacgtcagagccttaigtcatctaactcaactcctctgaagccggt 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1270 agacgctcactggccagtcgcccgccttattgtttctctgagacaattctctggagtaggg 1211

QY 307 agcgagaggtcgggttcggtgagctgag 336
    | | | | | | | | | | | | | | | | | |
DB 1210 taccattgtgactgagatgagggcggtagag 1181

RESULT 7
PCT-US94-08806-1/C
; Sequence 1, Application PC/TUS9408806
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; APPLICANT: 1201 Eastlake Avenue East
; APPLICANT: Seattle
; APPLICANT: WA
; APPLICANT: USA
; APPLICANT: 98102
; APPLICANT: University of Washington
; APPLICANT: Seattle
; APPLICANT: WA
; APPLICANT: 98195
; TITLE OF INVENTION: HEMATOPOIETIC PROTEIN AND MATERIALS AND
; TITLE OF INVENTION: METHODS FOR MAKING IT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
```

```
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/08806
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-12PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6600 ext 6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: 1081
FEATURE:
NAME/KEY: CDS
LOCATION: 105..1241
PCT-US94-08806-1
```

```
Query Match      6.9%; Score 28.4; DB 5; Length 1486;
Best Local Similarity 49.3%; Pred. No. 4.5;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 187 tggcgagaggtggtgcacacgtgcttagtacagagcgtctgctgagcagcaactga 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1330 tggacagatgacagcgtcttcacagccttcctggggagaccttgctcccgagaaactgc 1271

QY 247 tcatgcataggggacacgtcagagccttaigtcatctaactcaactcctctgaagccggt 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1270 agacgctcactggccagtcgcccgccttattgtttctctgagacaattctctggagtaggg 1211

QY 307 agcgagaggtcgggttcggtgagctgag 336
    | | | | | | | | | | | | | | | | | |
DB 1210 taccattgtgactgagatgagggcggtagag 1181

RESULT 8
PCT-US95-01775-1/C
; Sequence 1, Application PC/TUS9501775
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; APPLICANT: 1201 Eastlake Avenue East
; APPLICANT: Seattle
; APPLICANT: WA
; APPLICANT: USA
; APPLICANT: 98102
; APPLICANT: METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```



```

1: INFORMATION FOR SEQ ID NO: 1:
2: SEQUENCE CHARACTERISTICS:
3:     LENGTH: 1486 base pairs
4:     TYPE: nucleic acid
5:     STRANDEDNESS: double
6:     TOPOLOGY: linear
7:     MOLECULE TYPE: cDNA
8:     FEATURE:
9:         NAME/KEY: CDS
10:        LOCATION: 105..1241
11: PCT-US95-14932-1
12:
13: Query Match          6.9%: Score 28.4; DB 5; Length 1486;
14: Best Local Similarity 49.3%: Pred. No. 4.5;
15: Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0.
16:
17: Oy 187  tggccagagtggtgtgtccacgcgcgttaagtaacagagcctgtgtgcccagcaactga 246
18:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19: Db 1330  tggacacatgacagctgtcccttcacaccttccctggggaaacctgtgtcccgacagaaacctcc 1271
20:
21: Oy 247  tcatgcataaggagcagacactgcagccttatgtgtcattaaactcaccctgttaggcggt 306
22:      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
23: Db 1270  agagcctcactggtggccagtgcccccggcgctatgtttccttgagacaaattcttgggattgaggg 1211
24:
25: Oy 307  agcgacagaggtccggttcggtgtgagctggag 336
26:      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
27: Db 1210  taccattgtgactgcatgagtgagggcggtagag 1181
28:
29: RESULT 11
30: PCT-US95-16626-1/c
31: ; Sequence 1, Application PC/TUS9516626
32: ; GENERAL INFORMATION:
33: ; APPLICANT: ZymoGenetics, Inc.
34: ; APPLICANT: 1201 Eastlake Avenue East
35: ; APPLICANT: Seattle
36: ; APPLICANT: WA
37: ; APPLICANT: USA
38: ; APPLICANT: 98102
39: ; TITLE OF INVENTION: PURIFIED THROMBOPOIETIN AND METHOD OF
40: ; TITLE OF INVENTION: MAKING IT
41: ; NUMBER OF SEQUENCES: 24
42: ; CORRESPONDENCE ADDRESS:
43: ; ADDRESSEE: ZymoGenetics, Inc.
44: ; STREET: 1201 Eastlake Avenue East
45: ; CITY: Seattle
46: ; STATE: WA
47: ; COUNTRY: USA
48: ; ZIP: 98102
49: ; COMPUTER READABLE FORM:
50: ; MEDIUM TYPE: Floppy disk
51: ; COMPUTER: IBM PC compatible
52: ; OPERATING SYSTEM: PC-DOS/MS-DOS
53: ; SOFTWARE: PatentIn Release #1.0, Version #1.25
54: ; CURRENT APPLICATION DATA:
55: ; APPLICATION NUMBER: PCT/US95/16626
56: ; FILING DATE:
57: ; CLASSIFICATION:
58: ; ATTORNEY/AGENT INFORMATION:
59: ; NAME: Parker, Gary E
60: ; REGISTRATION NUMBER: 31,648
61: ; REFERENCE/DOCKET NUMBER: 94-11PC
62: ; TELECOMMUNICATION INFORMATION:
63: ; TELEPHONE: 206-442-6673
64: ; TELEFAX: 206-442-6678
65: ; INFORMATION FOR SEQ ID NO: 1:
66: ; SEQUENCE CHARACTERISTICS:
67: ; LENGTH: 1486 base pairs
68: ; TYPE: nucleic acid
69: ; STRANDEDNESS: double
70: ; TOPOLOGY: linear
71: ; MOLECULE TYPE: cDNA

```

```

: FEATURE:
: NAME/KEY: CDS
: LOCATION: 105..1241
PCT-US95-16626-1

Query Match          6.9%; Score 28.4; DB 5; Length 1486;
Best Local Similarity 49.3%; Pred. NO. 4.5;
Matches 74; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 187 tggcgaaggttgagctgcacacgttcctcagtaacagacgtctgctgcccagcaactga 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1330 tggacacagatgcagctgctcctcctcagccttccctggggaaagcttgcctcccgagaaagctgc 1271

Qy 247 tcatgcataagggcacaactgacagcgcctatgtgcatcaataacatccatcctgtaagccggt 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1270 agacgcctcactctggccagcagcgcctccggcgatgattttccttgagacaaatttccctggatgaggg 1211

Qy 307 agcgcagaaggtccggttccggtgtagcttgag 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 tacattgtgactgcatgtagagggcgctgagag 1181

RESULT 12
US-08-379-496-1
: Sequence 1, Application US/08379496
: Patent No. 5593833
: GENERAL INFORMATION:
: APPLICANT: MORRISON, Nigel A
: APPLICANT: EISMAN, John A
: APPLICANT: KELLY, Paul J
: TITLE OF INVENTION: Assessment of Trans-Acting Factors Allelic
: TITLE OF INVENTION: Variation
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rothwell, Fleg, Ernst & Kurz
: STREET: Suite 701-E, 555 13th Street.N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/379,496
: FILING DATE: 02-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ERNST, Barbara G
: REGISTRATION NUMBER: 30,377
: REFERENCE/DOCKET NUMBER: 1871-114
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 783-6040
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2169 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-379-496-1

Query Match          6.9%; Score 28.4; DB 1; Length 2169;
Best Local Similarity 51.6%; Pred. NO. 5.5;
Matches 65; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 64 tggaccaccatgaatgaagatctcccttcggtatgtctcccgaaagctaatccctca 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 963 tggagcgtcmaaggaatggagtgctctcttaccctccctgcacgccaagcagcttcttcatttc 1022

```



```
; Sequence 45, Application PC/TUS9311153
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: Novel Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11153
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 201..1211
; PCT-US93-11153-45

Query Match 6.8%; Score 28.2; DB 5; Length 1317;
Best Local Similarity 53.1%; Pred. No. 5;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 237 cagcaactgatacgtacgagcgacacgcgccttatgtgcatctaaactcactcct 296
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 747 CAACGACCCCAAGTCAATGGGTAGACGGTCACAGATATATCTGTCATCGCCTCAC 688
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 297 gtaagccggtagcgacaggtccggtcgtgagctcgagcgagcgaggtccctg 349
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 687 TGACGTTGGCAAGATGCGGAAATGATCAGCAGAGGAGGAGGATCCAG 635
```

Search completed: May 8, 2001, 15:14:05
Job time: 41063 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:21 ; Search time 3095.41 seconds
(without alignments)
1165.600 Million cell updates/sec

Title: US-09-619-643-2

Perfect score: 413
Sequence: 1 agtcgtgaatccatcgca.....ccccaggctggaaccggag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 segs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

```
1: gb_est1:**
2: gb_est2:**
3: gb_est3:**
4: gb_est4:**
5: gb_est5:**
6: gb_est6:**
7: gb_est7:**
8: gb_est8:**
9: gb_est9:**
10: gb_est10:**
11: gb_est11:**
12: gb_est12:**
13: gb_est13:**
14: gb_est14:**
15: gb_est15:**
16: gb_est16:**
17: gb_est17:**
18: gb_est18:**
19: gb_est19:**
20: gb_est20:**
21: gb_est21:**
22: gb_est22:**
23: gb_est23:**
24: gb_est24:**
25: gb_est25:**
26: gb_est26:**
27: gb_est27:**
28: gb_est28:**
29: gb_est29:**
30: gb_est30:**
31: gb_est31:**
32: gb_est32:**
33: gb_est33:**
34: gb_est34:**
35: gb_est35:**
36: gb_est36:**
37: gb_est37:**
38: gb_est38:**
39: gb_est39:**
40: gb_est40:**
41: em_est1:**
42: em_est2:**
43: em_est3:**
44: em_est4:**
45: em_est5:**
46: em_est6:**
47: em_est7:**
48: em_est8:**
49: em_est9:**
50: em_est10:**
51: em_est11:**
52: em_est12:**
53: em_est13:**
54: em_est14:**
55: em_est15:**
56: em_est16:**
57: em_est17:**
58: em_est18:**
59: em_est19:**
60: em_est20:**
61: em_est21:**
62: em_est22:**
63: em_est23:**
64: em_est24:**
65: em_est25:**
66: em_est26:**
67: em_est27:**
68: em_est28:**
69: em_est29:**
70: em_est30:**
71: em_est31:**
72: em_est32:**
73: em_est33:**
74: em_est34:**
75: em_est35:**
76: em_est36:**
77: em_est37:**
78: em_est38:**
79: em_est39:**
80: em_est40:**
81: em_est41:**
82: em_est42:**
83: em_est43:**
84: em_est44:**
85: em_est45:**
86: em_est46:**
87: em_est47:**
88: em_est48:**
89: em_est49:**
90: em_est50:**
91: em_est51:**
92: em_est52:**
93: em_est53:**
94: em_est54:**
95: em_est55:**
96: em_est56:**
97: em_est57:**
98: em_est58:**
99: em_est59:**
100: em_est60:**
101: em_est61:**
102: em_est62:**
103: em_est63:**
104: em_est64:**
105: em_est65:**
106: em_est66:**
107: em_est67:**
108: em_est68:**
109: em_est69:**
110: em_est70:**
111: em_est71:**
112: em_est72:**
113: em_est73:**
114: em_est74:**
115: em_est75:**
116: em_est76:**
```

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pln1:*
 191: em_gss_pln2:*
 192: em_gss_pro:*
 193: em_gss_rod1:*
 194: em_gss_rod2:*
 195: em_gss_rod3:*
 196: em_gss_rod4:*
 197: em_gss_rod5:*
 198: em_gss_vrl1:*
 199: em_gss_vrl2:*
 200: em_gss_vrl3:*
 201: gb_gss1:*
 202: gb_gss2:*
 203: gb_gss3:*
 204: gb_gss4:*
 205: gb_gss5:*
 206: gb_gss6:*
 207: gb_gss7:*
 208: gb_gss8:*
 209: gb_gss9:*
 210: gb_gss10:*
 211: gb_gss11:*
 212: gb_gss12:*
 213: gb_gss13:*
 214: gb_gss14:*
 215: gb_gss15:*
 216: gb_gss16:*
 217: gb_gss17:*
 218: gb_gss18:*
 219: gb_gss19:*
 220: gb_gss20:*
 221: gb_gss21:*
 222: gb_gss22:*
 223: gb_gss23:*
 224: gb_gss24:*
 225: gb_gss25:*
 226: gb_gss26:*
 227: gb_gss27:*
 228: gb_gss28:*
 229: gb_gss29:*
 230: gb_gss30:*
 231: gb_gss31:*
 232: gb_gss32:*
 233: gb_gss33:*
 234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198.8	48.1	619	166	BE358113
2	43.6	10.6	572	23	AL668136
3	39.2	9.5	415	145	BF199872
4	37	9.0	500	136	BE553481
5	35.4	8.6	664	23	AT650218
6	35.2	8.5	556	214	AQ965304
7	34.6	8.4	652	7	AA445084
8	34.6	8.4	844	229	CNS0052P
9	34.4	8.3	1101	229	CNS00FHE
10	33.8	8.2	482	140	BE860590
11	33.8	8.2	580	220	A2359473
12	33.4	8.1	931	230	CNS022KJ
13	33.2	8.0	769	136	BE515041
14	33	8.0	339	159	Z43278
15	33	8.0	380	116	AM488503
16	33	8.0	423	102	AI845446
17	33	8.0	603	2	AA109984
18	33	8.0	609	159	W96905

19	33	8.0	647	112	AM209908	
20	33	8.0	656	3	AA208975	
21	32.8	7.9	389	30	AA208933	
22	32.8	7.9	413	30	AA415171	
23	32.8	7.9	423	29	AA410984	
24	32.6	7.9	358	115	AA435290	
25	32.6	7.9	420	20	AI411532	
26	32.6	7.9	421	113	AM253244	
27	32.6	7.9	545	2	AA124272	
28	32.6	7.9	604	16	AI105270	
29	32.6	7.9	627	17	AI128249	
30	32.2	7.8	437	156	R98112	
31	32.2	7.8	479	107	AU098331	
32	32.2	7.8	1101	229	CNS0038X	
33	32.2	7.8	1101	231	CNS0406G	
34	32	7.7	308	111	AM125223	
35	32	7.7	375	175	C71138	
36	32	7.7	930	230	CNS03A3Z	
37	31.8	7.7	423	158	W66086	
38	31.8	7.7	583	147	BF346812	
39	31.8	7.7	865	216	AZ088424	
40	31.6	7.7	1350	144	BF179995	
41	31.6	7.7	339	116	AM526916	
42	31.6	7.7	545	162	BE095875	
43	31.6	7.7	558	216	AZ112251	
44	31.6	7.7	785	147	BF340077	
45	31.6	7.7	890	230	CNS0392B	

ALIGNMENTS

RESULT 1
LOCUS BE358113/c
DEFINITION DGI_24.A02.g1_A002 Dark Grown 1 (DGI) Sorghum bicolor CDNA, mRNA
ACCESSION BE358113
VERSION BE358113
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 619)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
An EST database from Sorghum: dark-grown seedlings
unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: POLYTMIX
High quality sequence start: 25
High quality sequence stop: 613
POLYA-No.

FEATURES
Source
1. 619
Location/Qualifiers

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGI)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT	155 a	134 c	148 g	182 t
ORIGIN				
Query Match	48.1%; Score 198.8; DB 166; Length 619;			
Best local Similarity	76.1%; Pred. No. 3.9e-52;			
Matches 303; Conservative	0; Mismatches 77; Indels 18; Gaps 4;			
Qy	1	agctcgaatcatcattgcaatgcatcgtctctcgtcggtctttttgtaactc-----	52	
Db	435	AGCTCGAAATACCAAGCATCGAATTCCTGGCGTTCTTTGTATCATCTCGGATCA	376	
Qy	53	cccaacaattatgagccaccatgaatgaagatccctccctcgtcgtacgtcccgaaaggc	112	
Db	375	CCACACAAATATGACACCAATGAGAGAGATCCCTTTCTGTATGCTCCAGAGAGGC	316	
Qy	113	taatccctcaatcaagcgcgtgaagctccagcatcaatgatatataatagga	172	
Db	315	TAATTAATTCATACAGCGCTGAGCTACAGACAGACATAGATATATACACTAGAC	256	
Qy	173	cttgcctgatccatgagcgaaggctgacacgcgtgccttagtacagacgtctgac	232	
Db	255	ACCTGATCGATCCATCGCTGCGAGAGTGAATTC-----GCCTGAAACAGATCCTGCGC	202	
Qy	233	tggccagcaactgatatcatgatataggcgacactgacagccttatgcttaactcca	292	
Db	201	TGGCGACCAATGATGATGATGAGAGGCGCACACTGCAACG--TTGATCTGCAATCCA	145	
Qy	293	tccttgagccgtagcagaggtccggttcggtgagcgtgagagaggttccttcac	352	
Db	144	TCCGCTAAGCCGCCATACAGAGCTCCGCTTGATTAATGAAGGACAGATTTCAGAAC	85	
Qy	353	cgatgccttgacgttcgacgagcgaatggaacctgagca	390	
Db	84	CGTGTGCTTGCG-ATCGCGTGCATGAGACTGAA	48	

RESULT 2
LOCUS A1668136
DEFINITION 60501SC08.x1 605 - Endosperm CDNA library from Schmidt lab Zea mays
ACCESSION A1668136
VERSION A1668136
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 572)
Maize ESTs from various CDNA libraries sequenced at Stanford University
unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605015 row: C column: 08.
Location/Qualifiers
1. 572
Location/Qualifiers

FEATURES
Source
1. 572
Location/Qualifiers
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm CDNA library from Schmidt lab"
/tissue="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5 (alpha)"
/note="Organ: Kernel; Vector: pMD-GAL4-2'; Site_1: EcoRI;

BASE COUNT	158 a	170 c	134 g	109 t	1 others
Query Match	10.6%	Score 43.6:	DB 23:	Length 572:	
Best Local Similarity	55.3%	Pred. No. 0.0056:			
Matches 194:	Conservative	0:	Mismatches 115:	Indels 42:	Gaps 4:
ORIGIN					
Db	8	aaatccatgtagatcgctctctcgcgcgcctctttgttcatccctccacaatatga	67		
Db	222	ACAATCCTGGCGCTCTTTTGTGTAGCTCTCGGCATACCAACAATATATGSA	281		
Qy	68	cc-----acatgaatgaatgatactccctcgatgct-----cccgaaagct	113		
Db	282	CTGGACACACCTGGAAACCGAAGATCCCTTTCCTGTAGCTCCCTCCCAAGGCT	341		
Qy	114	aatccctcaatcacagcgctgaagctccagatatactcaatgattatataatgagcc	173		
Db	342	AATCATTGATACACAGCGCTTACAGTACAGAAAGACACACGCTTACACACAC	401		
Qy	174	ctgtcgcataccatgycgcagggctgagctgccc-----acc	208		
Db	402	ACTAGACACACACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	461		
Qy	209	gtgcttagtacaagagctctcgtgctgacgaacatgataatgataagagcaact	265		
Db	462	GGCCTTAAACGAAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	521		
Qy	266	gcagcgctatgctgcatatcaactcatctctgtagcgcgtagcgagag	316		
Db	522	GCAGGCCGNGGTCTGCGAATGCTGATCTTAAACACGCGGACACACG	572		
RESULT	3				
LOCUS	BF199872/c				
DEFINITION	BF199872	415 bp	mRNA	EST	06-NOV-2000
ACCESSION	WHE0597_E08_J152E	Triticum monococcum shoot apex cDNA library			
VERSION	BF199872	Triticum monococcum cDNA clone WHE0597_E08_J15			
KEYWORDS	BF199872.1	GI:1114130			
SOURCE	EST.				
ORGANISM	Triticum monococcum.				
REFERENCE	Triticum monococcum.				
AUTHORS	Eukaryota: Viridiplantae; Embryophyta: Tracheophyta; Spermatophyta; Magnoliophyta: Liliopsida: Poales: Poaceae; Pooidae; Triticeae; Triticum.				
TITLE	1 (bases 1 to 415)				
JOURNAL	Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., Stomove, B. and Tong, J.C.				
COMMENT	The structure and function of the expressed portion of the wheat genomes - Shoot apex cDNA library from Triticum monococcum Unpublished (2000)				
FEATURES	Contact: Olin Anderson				
SOURCE	US Department of Agriculture, Agriculture Research Service, Pacific Northwest Area, Western Regional Research Center				
	800 Buchanan Street, Albany, CA 94710, USA				
	Tel: 5105595773				
	Fax: 5105595818				
	Email: oanderson@wv.usda.gov				
	Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20				
	Seq primer: Stragene T3 primer.				
	Location/Qualifiers				
	1..415				
	/organism="Triticum monococcum"				
	/cultivar="DV92"				
	/db_xref="taxon:4568"				
	/clone="WHE0597_E08_J15"				
	/clone_lib="Triticum monococcum shoot apex cDNA library"				
	/tissue_type="Vegetative shoot apex"				

Query Match	Score	DB	Length
Query Match	9.58	Score 39.2; DB 145; Length 415;	
Best Local Similarity	69.0%	Pred. No. 0.13;	
Matches 69; Conservative	0;	Mismatches 28; Indels 3; Gaps 1	
Db	213	ctagacagagcgtctgcgtgcgcagcaactatcatatgcatagggcagcactgcaagcc 272	
	210	CTAATACAACTACTCGGTTCGGGGCCACCTATCATGCGAGGAGCACACACTCGACGCC 151	
Qy	273	ttaatgctcataaaccatccatccctgtatggcggtagcgca 312	
Db	150	---TTGATCTTGCAAAACCATGCCGTAGCTGGGGGGGAA 114	
RESULT 4			
LOCUS	BE553481		
DEFINITION	u746g12.y1 NCI CGAP bp2 Mus musculus cDNA clone IMAGE:315382 5' similar to gb:M12783 PLATELET-DERIVED GROWTH FACTOR, B CHAIN PRECURSOR (HUMAN);, mRNA sequence.		
ACCESSION	BE553481		
VERSION	BE553481.1		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 500)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Other_ESRs: u746g12.x1 Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:1056138 Seq primer: -40RP from Gibco High quality sequence stop: 438.		
FEATURES			
SOURCE	Location/Qualifiers 1..500 /organism="Mus musculus" /strain="FVB/N-3" /db_xref="taxon:10090" /clone="IMAGE:315382" /clone_1lb="NCI_CGAP_Mam2" /tissue_type="tumor, biopsy sample" /dev_stage="5 months" /lab_host="DH10B" /note="Organ: mammary; Vector: PCWV-SPORN6; Site1: Saliv Site2: Noli; Cloned unidirectionally. Primer: Oligo dT library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"		
BASE COUNT	91 a 158 c 143 g 108 t		
ORIGIN			

[illegible]

ACCESSION	A0965304		
VERSION	A0965304.1		
KEYWORDS	GI:6793005		
SOURCE	GSS.		
ORGANISM	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 556)		
AUTHORS	Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Uterbach,T., Feldblum,T., Liang,F., Creasy,T. and Fraser,C.M.		
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atelidgr.org For additional information, see http://www.tigr.org/tdb/at/at.html Seq primer: TF		
FEATURES	Class: Shotgun.		
source	Location/Qualifiers		
	1..556		
	/organism="Arabidopsis thaliana"		
	/strain="Landsberg erecta"		
	/db_xref="taxon:3702"		
	/clone="LIRB45"		
	/clone_lib="LEBG"		
	/note="Organ: Leaf; Vector: pUC19/K; Total genomic DNA was sheared to 0.4-0.7 kbp before ligation."		
BASE COUNT	143 a	98 c	139 g 175 t 1 others
ORIGIN			
Query Match	8.5%	Score 35.2;	DB 214; Length 556;
Best Local Similarity	55.8%;	Pred. No. 2.5;	
Matches	67;	Conservative	0; Mismatches 53; Indels 0; Gaps 0;
OY	10	atccatgcacgcgtctctcgcgcctctttgttaccatccccaacaattatgacc	69
DB	28	ATCCATTGGATCATTTTGGAGTGCTTTTCTTTTGTCTTGCGGTTCAATACGTGACGTGG	87
OY	70	accatgaatgaagaatctcccttcggtcgtatgctcccgaaaggtcaatccctcaatag	129
DB	88	ACGATTACAGCAAACTCTCGCAGATCTGTGGAACTGAAAGGAATATCTCTGCACTCCGG	147
RESULT	7		
LOCUS	AA445084	652 bp	mRNA
DEFINITION	v959604.r1 Beddington mouse embryonic region Mus musculus cDNA clone IMAGE:655639 5' similar to gb:D28480 MCM3 HOMOLOG (HUMAN);, mRNA sequence.		
ACCESSION	AA445084		
VERSION	AA445084.1	GI:2157767	
KEYWORDS	EST.		
ORGANISM	house mouse.		
SOURCE	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 652)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.		
TITLE	The Mashu-HMT Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project Mashu-HMT Mouse EST Project		

FEATURES

Location/Qualifiers

1. 769

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3633289"

/clone_1lb="NIH-MGC-9"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

187 a 227 c 192 g 163 t

ORIGIN

Query Match 8.0%; Score 33.2; DB 136; Length 769;

Best Local Similarity 53.0%; Pred. No. 12;

Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 218 acagagcctctggcgcagcaactgacatgcatagagggcacactgcagccttatt 277

DB 241 ACAGGAGAGCCCGGACGGCCCTGCTCACATGAGAGATGGGACAGATGGGCAATATGCT 182

QY 278 ggtcattaaactccatctgtagcgcgtagcagagatcgcgttgcgtgtagctggag 337

DB 181 GGCTGTAGTACACCCCTCTGTAGGGGGGTCTGGCCAGGAGGTCTAGGAGGTGATGCCGGG 122

QY 338 gcagagttcctgca 351

DB 121 TCAGCCGCTACTGCA 108

RESULT 14

LOCUS

243278 339 bp mRNA EST 11-NOV-1994

DEFINITION HSC17H121 normalized infant brain cDNA Homo sapiens cDNA clone

ACCESSION 243278

VERSION 243278.1 GI:570864

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 339)

Auffray, C., Benar, G., Bois, F., Bouchier, C., da Silva, C., Devignes

M.D., Duprat, S., Houligate, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,

Sebastiani-Kabaktchis, C. and Tessier, A.

IMAGE: molecular integration of the analysis of the human genome

and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

JOURNAL

MEDLINE

COMMENT

Contact: Genethon

Genexpress-Genethon

Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpress@genethon.fr

Single read.

Genexpress library idt: C: Genexpress_sequence_idt: ylc-17h12

Seq primer: (-21)M13_universal.

/clone_1lb="normalized infant brain cDNA"

/sex="Female"

/tissue_type="total brain"

/dev_stage="3 months old"

/note="Organ: brain; Vector: latmid BA; Site_1: HindIII;

Site_2: NotI; sex=Female; dev_stage=3 months old;

isolate-muscular atrophy patient; tissue_type=total brain

; total mRNA was oligo-(dT) primed and directionally

cloned 5' -> 3' into the HindIII -> NotI sites of the

latmid BA vector. Clone library from B. Soares, Psychiatry

Dept. Columbia University, USA. Normalization_method:

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

Bento Soares, P.N.A.S. in press"

```

/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker: Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldi, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=hypothalamus
TAG_SEQ=CGGTA"

```

```

BASE COUNT      70 a      105 c      130 g      75 t
ORIGIN

```

```

Query Match      8.0%: Score 33; DB 116; Length 380;
Best Local Similarity 57.1%: Pred No. 11;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

```

QY 176 tgcctgcatccatgctgcaggtgctgctgacacgtgctagtcagagcactctggtgtg 235
    ||| ||||| || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 189 TGCAGGATCCACAGCCAGCGCTGGTGGGGCTGGGAGTGGGGCTGCACAG 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 236 ccagcacactgcatgcatagggcacactgcagccttatgtgt 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 249 CCAAGAAAGTAGATCCAGAGGGGAGCTCAGGAGGCCCAAGGCT 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Search completed: May 8, 2001, 05:34:26
Job time: 9035 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:48:11 ; Search time 2791.02 Seconds
(Without alignments)
1928.674 Million cell updates/sec

Title: US-09-619-643-3

Perfect score: 365

Sequence: 1 cgatggcgcgattgtgttc.....ctaatactccctacattcta 365

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htg_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vl:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vl1:*

59: gb_vl2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_ro1:*

95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	167.8	46.0	1596	15	ZMSERTSYN	Y13053 Zea mays mR
2	43.8	12.0	2954	94	GPIESTRSULF	L11117 Guinea pig
3	37.6	10.3	196955	74	AC068606	AC068606 Mus muscu
4	37.6	10.3	293219	76	AC079577	AC079577 Mus muscu
5	37.6	10.3	310148	75	AC073729	AC073729 Mus muscu
6	37.4	10.2	113693	92	HSRDI1C16	AL049814 Human DNA
7	37.4	10.2	163694	69	AC024499	AC024499 Homo sapi
8	37.4	10.2	164055	81	AL450348	AL450348 Homo sapi
9	37	10.1	156838	66	AC020672	AC020672 Homo sapi
10	36.6	10.0	169348	86	AC008699	AC008699 Homo sapi
11	36	9.9	110000	75	AC073794_2	Continuation (3 of

C	12	35.6	9.8	180672	87	AC011451	AC011451	Homo	sapl
C	13	35.2	9.6	121673	75	AC073949	AC073949	Homo	sapl
C	14	35.2	9.6	202029	87	AC010087	AC010087	Homo	sapl
C	15	35.2	9.6	205818	77	AC087062	AC087062	Mus	sapl
C	16	35	9.6	3027	5	AE217633	AE217633	Drosophila	
C	17	35	9.6	99228	89	AL133376	AL133376	Human	DNA
C	18	35	9.6	149143	62	AC011591	AC011591	Homo	sapl
C	19	35	9.6	157277	67	AC022048	AC022048	Homo	sapl
C	20	35	9.6	193622	61	AC009659	AC009659	Homo	sapl
C	21	35	9.6	196955	74	AC068606	AC068606	Mus	sapl
C	22	35	9.6	209575	75	AC073739	AC073739	Mus	sapl
C	23	34.8	9.5	141643	78	AE121897	AE121897	Homo	sapl
C	24	34.8	9.5	185982	75	AC073231	AC073231	Homo	sapl
C	25	34.8	9.5	340000	91	HS21C080	HS21C080	Homo	sapl
C	26	34.4	9.4	58880	70	AC052855	AC052855	Homo	sapl
C	27	34.4	9.4	59205	77	AC084254	AC084254	Homo	sapl
C	28	34.4	9.4	66460	77	AC083840	AC083840	Homo	sapl
C	29	34.4	9.4	115833	82	AP001942	AP001942	Homo	sapl
C	30	34.4	9.4	179681	63	AC001376	AC001376	Homo	sapl
C	31	34.4	9.4	198895	82	AP001337	AP001337	Homo	sapl
C	32	34.4	9.4	204623	86	AP001338	AP001338	Homo	sapl
C	33	34	9.3	127811	60	AC008375	AC008375	Homo	sapl
C	34	34	9.3	152423	88	AC087432	AC087432	Homo	sapl
C	35	34	9.3	171959	88	AC087427	AC087427	Homo	sapl
C	36	34	9.3	173659	65	AC077110	AC077110	Homo	sapl
C	37	34	9.3	189483	86	AC008733	AC008733	Homo	sapl
C	38	34	9.3	201981	75	AC073640	AC073640	Homo	sapl
C	39	34	9.3	202140	77	AC087097	AC087097	Mus	sapl
C	40	33.8	9.3	136150	13	AP002485	AP002485	Oryza	satl
C	41	33.8	9.3	188138	65	AC001866	AC001866	Homo	sapl
C	42	33.6	9.2	60159	89	AL355504	AL355504	Human	DNA
C	43	33.6	9.2	135849	88	AF141556	AF141556	Human	DNA
C	44	33.6	9.2	144165	85	AC002127	AC002127	Human	BAC
C	45	33.6	9.2	144165	85	AC002127	AC002127	Human	BAC

ALIGNMENTS

AC011451	Human	sapB
AC079949	Human	sapB
AC010087	Human	sapB
AC087062	Mus musculus	
AA121763	Drosophila	
AF113337	Human	DNA
AC011591	Human	sapB
AC022048	Human	sapB
AC096659	Human	sapB
AC066606	Mus musculus	
AC073793	Mus musculus	
AF121897	Human	sapB
AC075221	Human	sapB
AL163280	Human	sapB
AC025285	Human	sapB
AC084254	Human	sapB
AC083840	Human	sapB
AP001942	Human	sapB
AC013785	Human	sapB
AP001337	Human	sapB
AP001338	Human	sapB
AC020928	Human	sapB
AC068375	Human	sapB
AC087432	Human	sapB
AC087427	Human	sapB
AC071110	Human	sapB
AC008733	Human	sapB
AC073640	Human	sapB
AC087097	Mus musculus	
AP003485	Oryza sativa	
AC018866	Human	sapB
AL155504	Human	DNA
AF131456	Human	sapB
AC002127	Human	BAC

```

RESULT 1
ZMSETSTYN
LOCUS ZMSETSTYN 1596 bp mRNA PLN 24-JUL-1998
DEFINITION zea mays mRNA for seryl-tRNA synthetase.
ACCESSION Y13053
VERSION Y13053.1 GI:3355716
KEYWORDS seryl-tRNA synthetase.
SOURCE zea mays.
ORGANISM zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Zea.
REFERENCE 1 (bases 1 to 1596)
Rokov,J., Soll,D. and Weygand-Durasevic,I.
Maize mitochondrial seryl-tRNA synthetase recognizes Escherichia
coli tRNA(Ser) in vivo and in vitro
Plant Mol. Biol. 38 (3), 497-502 (1998)
98418629
REFERENCE 2 (bases 1 to 1596)
Weygand-Durasevic,I.
Direct Submission
Submitted (07-MAY-1997) I. Weygand-Durasevic, Faculty of Science,
University of Zagreb, Department of Chemistry, Strossmayerov trg
14, 10000 Zagreb, CROATIA
Revised by [3]
3 (bases 1 to 1596)
Weygand-Durasevic,I.
Direct Submission
Submitted (05-JAN-1998) I. Weygand-Durasevic, Faculty of Science,
University of Zagreb, Department of Chemistry, Strossmayerov trg
14, 10000 Zagreb, CROATIA
Location/Qualifiers
1..1596
/organism="Zea mays"
FEATURES
source

```

```

/db_xref="taxon:4577"
/cell_line="LM195"
<1..1470
/EC_number="6.1.1.11"
/note="putative mitochondrial; evidence for activity of
enzyme is experimental, evidence for localization in the
cell is not experimental"
/codon_start=1
/product="seryl-tRNA synthetase"
/protein_id="CAA73496.1"
/db_xref="GI:3355717"
/db_xref="SPTRMBL:082110"
/translation="RTITRSLVRRPAPHLLSSASAAATAVELDTNGDSAGGAGY
RPOKKAIIDRKMINDRDVAADNTRKNSLSTANLELDVLELDQILTLQKVERLRAENR
AVAKKAKLIDPSYRQALVEGKNLKGSLALAEEDLVQLDLQLEAQS.IPTTHDPV
PVGGESVLRKEVSGRSRNFPAIKDLQLGKELDLPDAAEVSSEKRYLKNVAV
LLEALVYMAISEVSKGFTPLIPETIVRSVVEKCGFSKGNQVYSIEDSQCLIG
TAELPVGSHMDSILAEASALPIKVAVSHCFRPMAGAAATGVLVHVOFSKEMFI
PCRPESGKHMEELITTEPDYASICLAFPTLMATGDLCAPAYRKRDITAMPLER
YGEISSNSCTDIOSRRLGTRIRPSPEPSTSTKKKCAASGPTQFVHTLMTAAVAV
PRLIVCTLENNQDGSIIVPEPLRPYMGSLLELSPFK"

BASE COUNT      420 a      391 c      409 g      376 t
ORIGIN

Query Match      46.0%; Score 167.8; DB 15; Length 1596;
Best Local Similarity 95.3%; Pred. No. 1.3e-41;
Matches 184; Conservative 0; Mismatches 7; Indels 2; Gaps 1;

Oy      1 cgaatggtcgaattggtgtccagagacacacataaccatataatggtgacagagttgct 60
        |||||||
Db      1392 CGATGGGTGCGATGTGGTGTCCAGACCCACTAAGACCCCTATATGGGTGACACTAGACTTCT 1451

Oy      61 ctgggcaaatcaatcgaatgaccctctcactcgaatccgcctgtgtgct--tgtagactaga 118
        |||||||
Db      1452 CTCGCCAAATTCAGATGACCCCTCTACCGTCAACCGCTGTGTGCTGTGTGGACTTAGA 1511

Oy      119 actgtctgctggaatgtaaaagttaaggaagtctctctcactacaaatatacactact 178
        |||||||
Db      1512 ACTGCTGTGGTGAATGACAAAGTTAAAGGAAGTCAATTTCTCACTACAAATATACACTACT 1571

Oy      179 gaatttcgtgatt 191
        |||||||
Db      1572 GAGTTTCGTGATT 1584

RESULT 2
GPIESTSULF      2954 bp      DNA      ROD      24-AUG-1993
LOCUS      GPIESTSULF      2954 bp      DNA      ROD      24-AUG-1993
DEFINITION      Guinea pig estrone sulfotransferase gene.
ACCESSION      L11117
VERSION      L11117.1 GI:385214
KEYWORDS      estrone sulfotransferase.
SOURCE      Cavia porcellus DNA.
ORGANISM      Cavia porcellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Myiuricognathi; Cavillidae; Cavia.
REFERENCE      1 (bases 1 to 2954)
AUTHORS      Komatsu, K., Oeda, T. and Strodt, C. A.
TITLE      Cloning and sequence analysis of the 5' flanking region of the
estrone sulfotransferase gene: Steroid response elements and
cell-specific nuclear DNA-binding proteins
JOURNAL      Biochem. Biophys. Res. Commun. 194, 1297-1304 (1993)
MEDLINE      93356804
FEATURES
     source
         1..2954
         /organism="Cavia porcellus"
         /db_xref="taxon:10141"
         457..461
         1908..1912
         2435..2440
         2859..2863
         TATA_signal
         2889..2934
         tRNA

```



```
misc_feature      /note="assembly-fragment"
                  94749..110647
                  /note="assembly-fragment"
misc_feature      /note="assembly-fragment"
                  110748..122566
                  /note="assembly-fragment"
                  clone_end:sp6
                  vector_side:right"
misc_feature      122667..129447
                  /note="assembly-fragment"
misc_feature      129548..148164
                  /note="assembly-fragment"
misc_feature      148265..170684
                  /note="assembly-fragment"
                  clone_end:77
                  vector_side:right"
misc_feature      /note="assembly-fragment"
                  170785..196955
BASE COUNT      58538 a 38234 c 38014 g 60447 t 1722 others
ORIGIN
Query Match      10.3%; Score 37.6; DB 74; Length 196955;
Best Local Similarity 50.0%; Pred. No. 0.61;
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 157 ctccacacaataatacactcgaattcgtgattggggctctcttccataagcagtc 216
    ||||| ||| ||| ||| ||| ||||| ||| |||
DB 91583 CTCACGTCATGATGATGCTGCTGAGATCCTGCTTCTGACACCATCATG 91524
QY 217 gactactcgtcgtatcgtcagatagatactatgctgagcgtactcgtatcagataga 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91523 TTCTATCCCTCATCAGGAATTTCTCTATTATTCCTGCTGCTGCTGCTGCTGGAATCC 91464
QY 277 ctacaaagctatctgcatgcatagaccctcctatgctcccaagaatcagtgacattt 336
    || ||| ||| ||| ||| ||||| ||| ||| |||
DB 91463 TGCAGCTTGGATTCACAGACACGCGCCCTATTATGCTTCCCAACTGTTACAGATGCTA 91404
QY 337 aaaagcat 344
    ||| |||
DB 91403 TAAATTTT 91396

RESULT 4
AC079577 293219 bp DNA 02-SEP-2000
LOCUS Mus musculus clone Rp23-88J13, WORKING DRAFT SEQUENCE, 38 unordered
DEFINITION pieces.
ACCESSION AC079577
VERSION AC079577.1 GI:9964942
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 293219)
DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 293219)
REFERENCE Direct Submission
TITLE Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Mainut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1770893
Center clone name: RPCI-23_88J13
-----
Summary Statistics
```

```
Consensus quality: 263637 bases at least Q40
Consensus quality: 282586 bases at least Q20
Consensus quality: 286018 bases at least Q20
Estimated insert size: 186190; agarose-fp estimation
Estimated insert size: 289519; sum-of-contigs estimation
Quality coverage: 11.83 in Q20 bases; agarose-fp estimation
Quality coverage: 7.61 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1004: contig of 1004 bp in length
* 1005 1104: gap of unknown length
* 1105 2491: contig of 1387 bp in length
* 2492 2591: gap of unknown length
* 2592 3911: contig of 1320 bp in length
* 3912 4011: gap of unknown length
* 4012 5675: contig of 1664 bp in length
* 5676 5775: gap of unknown length
* 5776 7313: contig of 1538 bp in length
* 7314 7414: gap of unknown length
* 7414 8710: contig of 1297 bp in length
* 8711 8810: gap of unknown length
* 8811 10844: contig of 2034 bp in length
* 10845 10944: gap of unknown length
* 10945 12118: contig of 1174 bp in length
* 12119 12218: gap of unknown length
* 12219 14988: contig of 2770 bp in length
* 14988 15088: gap of unknown length
* 15089 16648: contig of 1560 bp in length
* 16649 16748: gap of unknown length
* 16749 19230: contig of 2482 bp in length
* 19231 19330: gap of unknown length
* 19331 22823: contig of 3493 bp in length
* 22824 22923: gap of unknown length
* 22924 26288: contig of 3365 bp in length
* 26289 26388: gap of unknown length
* 26389 29138: contig of 2750 bp in length
* 29139 29238: gap of unknown length
* 29239 31358: contig of 2120 bp in length
* 31359 31458: gap of unknown length
* 31459 33157: contig of 1699 bp in length
* 33158 33257: gap of unknown length
* 33258 35757: contig of 2500 bp in length
* 35758 35857: gap of unknown length
* 35858 38598: contig of 2741 bp in length
* 38599 38698: gap of unknown length
* 38699 41132: contig of 2434 bp in length
* 41133 41233: gap of unknown length
* 41233 45193: contig of 3961 bp in length
* 45194 45293: gap of unknown length
* 45294 49247: contig of 3924 bp in length
* 49248 49347: gap of unknown length
* 49348 54584: contig of 5237 bp in length
* 54585 54684: gap of unknown length
* 54685 58864: contig of 4180 bp in length
* 58865 58964: gap of unknown length
* 58965 64047: contig of 5083 bp in length
* 64048 64147: gap of unknown length
* 64148 71817: contig of 7670 bp in length
* 71818 71917: gap of unknown length
* 71918 81357: contig of 9440 bp in length
* 81358 81457: gap of unknown length
* 81458 89446: contig of 7969 bp in length
* 89447 89546: gap of unknown length
* 89547 100218: contig of 10672 bp in length
* 100219 100318: gap of unknown length
* 100319 108923: contig of 8605 bp in length
* 108924 109023: gap of unknown length
```

```
* 109024 117643: contig of 8620 bp in length
* 117644 117743: gap of unknown length
* 117744 128332: contig of 10589 bp in length
* 128333 128432: gap of unknown length
* 128433 142388: contig of 13956 bp in length
* 142389 142488: gap of unknown length
* 142489 165070: contig of 22582 bp in length
* 165071 165170: gap of unknown length
* 165171 186252: contig of 21082 bp in length
* 186253 186352: gap of unknown length
* 186353 205760: contig of 19408 bp in length
* 205761 205860: gap of unknown length
* 225914 225914: contig of 20054 bp in length
* 225915 226014: gap of unknown length
* 226015 252425: contig of 26411 bp in length
* 252426 252525: gap of unknown length
* 252526 293219: contig of 40694 bp in length.
Location/Qualifiers
1. 293219
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-88J13"
/clone_id="RP23-88J13"
BASE COUNT 88217 a 55579 c 56181 g 89524 t 3718 others
ORIGIN
Query Match 10.3% Score 37.6; DB 76; Length 293219;
Best Local Similarity 50.0%; Pred. No. 0.62;
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
OY 157 ctccataaatactactactgaatttcgtagtgagggctctgttcataagcagtc 216
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110112 CTCACGTGCAATGAGCATGTTGGTCTGTCAGAGATCTGGCTTGTGACACACATCAATG 110171
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 217 gactactctgcatcgtatcgtacatatactatgctgctactctgcatcgtacagca 276
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110172 TTCTATCCTCATGACGATTCCTTCTATTAATCTCTGCTGTGCTGTCCTGAGATCC 110231
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 277 ctcaaaactatgcatgcatgcatgagccctctatggttcccaagaatcagatgcat 336
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110232 TGCACGTTGGATGACGACGACTGGCCCTATTATGCTCCACACTGTTACAGATGCTA 110291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 337 aaagat 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110292 TAAATTTT 110299
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 5
AC073729/c 310148 bp DNA HTG 29-JUN-2000
LOCUS MUS MUSCULUS clone RP23-240L2, WORKING DRAFT SEQUENCE, 61 unordered
DEFINITION pieces.
AC073729 AC073729.1 GI:8810346
VERSION HTG: HTGS PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 310148)
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Mouse
REFERENCE 2 (bases 1 to 310148)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
```

```
-----
Project Information
Center Project Name: 1829298
Center clone name: RP23-240L2
Summary Statistics
Consensus quality: 263120 bases at least Q40
Consensus quality: 286215 bases at least Q30
Estimated insert size: 226920; agarose-1p estimation
Estimated insert size: 304148; sum-of-contigs estimation
Quality coverage: 9.47 in Q20 bases; agarose-1p estimation
Quality coverage: 7.07 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 61 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1048: contig of 1048 bp in length
1049 1148: gap of unknown length
1149 2407: contig of 1259 bp in length
2408 2507: gap of unknown length
2508 3536: contig of 1029 bp in length
3537 3636: gap of unknown length
3637 4872: contig of 1236 bp in length
4873 4972: gap of unknown length
4973 6083: contig of 1111 bp in length
6084 6183: gap of unknown length
6184 7814: contig of 1631 bp in length
7815 7914: gap of unknown length
7915 9306: contig of 1392 bp in length
9307 9406: gap of unknown length
9407 10509: contig of 1103 bp in length
10510 10609: gap of unknown length
10610 12157: contig of 1548 bp in length
12158 12257: gap of unknown length
12258 13387: contig of 1130 bp in length
13388 13487: gap of unknown length
13488 14559: contig of 1072 bp in length
14560 14659: gap of unknown length
14660 15855: contig of 1196 bp in length
15856 15955: gap of unknown length
15956 17518: contig of 1563 bp in length
17519 17618: gap of unknown length
17619 18750: contig of 1132 bp in length
18751 18850: gap of unknown length
18851 19959: contig of 1109 bp in length
19960 20059: gap of unknown length
20060 21407: contig of 1348 bp in length
21408 21507: gap of unknown length
21508 22835: contig of 1328 bp in length
22836 22935: gap of unknown length
22936 24081: contig of 1146 bp in length
24082 24181: gap of unknown length
24182 25679: contig of 1498 bp in length
25680 25780: gap of unknown length
25781 27045: contig of 1266 bp in length
27046 27145: gap of unknown length
27146 28617: contig of 1472 bp in length
28618 28717: gap of unknown length
28718 30530: contig of 1813 bp in length
30531 30631: gap of unknown length
30632 32151: contig of 1521 bp in length
32152 32251: gap of unknown length
32252 33981: contig of 1730 bp in length
33982 34081: gap of unknown length
34082 35262: contig of 1181 bp in length
35263 35362: gap of unknown length
35363 36660: contig of 1298 bp in length
36661 36760: gap of unknown length
```

FEATURES	SOURCE
36761	38195: contig of 135 bp in length
38196	38295: gap of unknown length
38296	40157: contig of 1862 bp in length
40158	40257: gap of unknown length
40258	42202: contig of 1945 bp in length
42203	42302: gap of unknown length
42303	44169: contig of 1867 bp in length
44170	44269: gap of unknown length
44270	44594: contig of 1225 bp in length
45495	45594: gap of unknown length
45595	47394: contig of 1800 bp in length
47395	47994: gap of unknown length
47995	49250: contig of 1756 bp in length
49251	49350: gap of unknown length
49351	51667: contig of 2317 bp in length
51668	51767: gap of unknown length
51768	53871: contig of 2104 bp in length
53872	53971: gap of unknown length
53972	55615: contig of 2544 bp in length
55616	56615: gap of unknown length
56616	60638: contig of 4023 bp in length
60639	60738: gap of unknown length
60739	62829: contig of 2091 bp in length
62830	62929: gap of unknown length
62930	64990: contig of 2061 bp in length
64991	65990: gap of unknown length
65991	67450: contig of 2360 bp in length
67451	67550: gap of unknown length
67551	70267: contig of 2717 bp in length
70268	70367: gap of unknown length
70368	72558: contig of 2191 bp in length
72559	72658: gap of unknown length
72659	76237: contig of 3579 bp in length
76238	76338: gap of unknown length
76338	81049: contig of 4712 bp in length
81050	81149: gap of unknown length
81150	84881: contig of 3732 bp in length
84882	84981: gap of unknown length
84982	88592: contig of 3611 bp in length
88593	88692: gap of unknown length
88693	93331: contig of 4639 bp in length
93332	93431: gap of unknown length
93432	102487: contig of 9056 bp in length
102488	102587: gap of unknown length
102588	108833: contig of 6546 bp in length
108834	108933: gap of unknown length
108934	117761: contig of 8828 bp in length
117762	117861: gap of unknown length
117862	126786: contig of 8925 bp in length
126787	126886: gap of unknown length
126887	136293: contig of 11407 bp in length
136294	136893: gap of unknown length
138394	149317: contig of 10924 bp in length
149318	149417: gap of unknown length
149418	162715: contig of 13298 bp in length
162716	162815: gap of unknown length
162816	179484: contig of 16666 bp in length
179485	179584: gap of unknown length
179585	196881: contig of 18897 bp in length
196882	198581: gap of unknown length
198582	215461: contig of 16880 bp in length
215462	215561: gap of unknown length
215562	235666: contig of 20105 bp in length
235667	235766: gap of unknown length
235767	254761: contig of 16995 bp in length
254762	254861: gap of unknown length
254862	282144: contig of 27283 bp in length
282145	282244: gap of unknown length
282245	310148: contig of 27904 bp in length
	Location/Qualifiers
	1.. 310148
	"/organism="Mus musculus"
	"/db_xref="taxon:10090"

BASE COUNT	93475	a	59392	c	60101	g	91157	t	6023	others
ORIGIN	/clone="RP23-240L2"									
	/clone=lpB="RPC1 mouse BAC library 23"									
Query Match	10.3%	Score 37.6;	DB 75;	Length 310148;						
Best Local Similarity	50.0%	Pred. No. 0.62;								
Matches 94:	Conservative	0;	Mismatches	94;	Indels	0;	Gaps	0;		
OY 157	ctcacataaataacatactgaattcgtatgctgagggctctgttctcataaagcagc	216								
Db 174280	CTCACTGCAGAAAGACATATGGTCTGCTACAGAACTCTGCTGTGCACACCATCATG	174221								
OY 217	gacactctgctgcatctgcatgcatgcatatatacctgctgcatctgcatgcatgca	276								
Db 174220	TTCTATCTCTATCAGGAATTTCTTCCATTATTTGCTGCTTCTGCTGTCTCGAGATCC	174161								
OY 277	ctacaagaactatgcatgcatgcatgagccctctatgctccaaagaatcgaatgacatt	336								
Db 174160	TGCAGCTTTGATGATCGACGACACTGGCCCTATTATATGCTTCCAACTGTTACAGATGCTA	174101								
OY 337	aaagatc 344									
Db 174100	TAAATTTT 174093									
RESULT 6										
HSJ1C16/c										
LOCUS	HSJ1C16	113693	bp	DNA	PRI	11-FEB-2000				
DEFINITION	Human DNA sequence from clone RPI-1C16 on chromosome 6p12.2-21.1									
ACCESSION	AL049814									
VERSION	AL049814.6 GI:5262385									
KEYWORDS	HTG.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;									
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.									
TITLE	1 (bases 1 to 113693)									
JOURNAL	Harley, J.									
COMMENT	Direct Submission Submitted (11-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquereyes@sanger.ac.uk requests: clonequests@sanger.ac.uk On Jun 29, 1999 this sequence version replaced g1:5019497. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP. Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/chr6 RPI-1C16 is from the library RPC1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 Location/Qualifiers									
FEATURES										

```
source
1. .113693
/organism="Homo sapiens"
/db xref="taxon:9606"
/chromosome="6"
/map="p12.2-21.1"
/clone_lib="RPC1-1"
/clone="RPC1-1C16"
1. .197
repeat_region
/note="AluJb repeat: matches 50. .237 of consensus"
2493. .2572
/note="2 copies 40 mer 88% conserved"
2574. .2778
repeat_region
/note="MIR repeat: matches 5. .234 of consensus"
2885. .3174
/note="L2 repeat: matches 2380. .2667 of consensus"
3281. .3578
/note="AluJ repeat: matches 1. .296 of consensus"
3585. .3950
/note="MIRJ repeat: matches 32. .418 of consensus"
4047. .4178
/note="66 copies 2 mer aa 59% conserved"
4388. .4613
repeat_region
/note="MIR repeat: matches 3. .243 of consensus"
5395. .5440
/note="Charliel repeat: matches 96. .141 of consensus"
5458. .5786
/note="AluSx repeat: matches 1. .299 of consensus"
5793. .6088
/note="AluJ repeat: matches 1. .296 of consensus"
6089. .6600
/note="Charliel repeat: matches 2221. .2759 of consensus"
6615. .7458
repeat_region
/note="L1MC1 repeat: matches 5449. .6332 of consensus"
7459. .8553
/note="HSMAR1 repeat: matches 197. .1287 of consensus"
8554. .8844
/note="L1PBI repeat: matches 5874. .6151 of consensus"
8871. .9076
/note="HSMAR1 repeat: matches 1. .207 of consensus"
9077. .9539
/note="L1MC1 repeat: matches 4981. .5401 of consensus"
9588. .10321
/note="L1MC1 repeat: matches 4477. .5291 of consensus"
10322. .10617
/note="AluJ repeat: matches 1. .295 of consensus"
10618. .11130
/note="L1MC1 repeat: matches 3988. .4477 of consensus"
11168. .11463
/note="AluSx repeat: matches 13. .307 of consensus"
11468. .12743
/note="L1MC repeat: matches 253. .1553 of consensus"
complement(12744. .12984)
/note="match: GSS: Em:AQ232631"
13143. .13243
/note="MIR repeat: matches 161. .262 of consensus"
13315. .13479
/note="MIR repeat: matches 8. .191 of consensus"
14078. .14130
/note="AluS repeat: matches 5. .57 of consensus"
14135. .14211
/note="MIRJ1 repeat: matches 288. .364 of consensus"
14300. .14594
/note="AluSx repeat: matches 1. .294 of consensus"
14845. .14908
/note="32 copies 2 mer aa 67% conserved"
16591. .17352
/note="L2 repeat: matches 1906. .2709 of consensus"
17685. .17962
/note="AluSx repeat: matches 3. .295 of consensus"
19262. .19560
/note="AluSx repeat: matches 1. .293 of consensus"
19600. .19895
/note="AluSx repeat: matches 1. .303 of consensus"

repeat_region
20445. .20530
/note="L2 repeat: matches 2532. .2602 of consensus"
20531. .20585
/note="MIR repeat: matches 211. .262 of consensus"
21171. .21298
/note="MIR repeat: matches 85. .253 of consensus"
24218. .24516
/note="AluSx repeat: matches 1. .299 of consensus"
25127. .25181
/note="MIR repeat: matches 115. .172 of consensus"
25356. .25486
/note="FLAM_C repeat: matches 1. .133 of consensus"
26495. .26509
/note="MER51-internal repeat: matches 7424. .7438 of consensus"
26510. .26602
/note="MER57-internal repeat: matches 7151. .7242 of consensus"
26721. .27037
/note="AluSx repeat: matches 1. .300 of consensus"
27282. .27686
/note="LIR37A repeat: matches 21. .426 of consensus"
27984. .28162
/note="HAL1 repeat: matches 757. .935 of consensus"
28163. .28248
/note="match: GSS: Em:AQ482783 Em:AQ508240"
28249. .28378
/note="FLAM_A repeat: matches 1. .130 of consensus"
28392. .28585
/note="97 copies 2 mer tt 68% conserved"
28394. .28593
/note="5 copies 40 mer 72% conserved"
28595. .28895
/note="AluSx repeat: matches 1. .306 of consensus"
28901. .29201
/note="AluSg repeat: matches 1. .299 of consensus"
29218. .29654
/note="HAL1 repeat: matches 202. .686 of consensus"
29774. .30040
/note="L1MC repeat: matches 269. .538 of consensus"
30041. .30068
/note="14 copies 2 mer gt 96% conserved"
30169. .30814
/note="L1M4 repeat: matches 3515. .4173 of consensus"
31130. .31191
/note="MIR repeat: matches 7. .67 of consensus"
31350. .31393
/note="L2 repeat: matches 2688. .2730 of consensus"
31428. .31491
/note="MIR repeat: matches 73. .136 of consensus"
31571. .31760
/note="MER5A repeat: matches 1. .189 of consensus"
34057. .34124
/note="MIR repeat: matches 147. .202 of consensus"
34125. .34444
/note="AluSg repeat: matches 1. .313 of consensus"
34445. .34537
/note="MIR repeat: matches 14. .147 of consensus"
35619. .35764
/note="MIR repeat: matches 100. .256 of consensus"
36247. .36563
/note="AluJb repeat: matches 1. .310 of consensus"
37088. .37370
/note="AluSx repeat: matches 1. .284 of consensus"
37585. .37706
/note="FLAM_A repeat: matches 1. .122 of consensus"
38167. .38337
/note="LIR16A repeat: matches 18. .186 of consensus"
38913. .38459
/note="HERV16 repeat: matches 1609. .2236 of consensus"
39497. .39946
/note="HERV16 repeat: matches 1112. .1553 of consensus"
39947. .40249
```


	misc_feature	37489..50434	/note="assembly_fragment"
	misc_feature	50535..62435	/note="assembly_fragment"
	misc_feature	clone_end:SP6	
	misc_feature	vector_side:left"	
	misc_feature	62536..78528	/note="assembly_fragment"
	misc_feature	78629..112818	/note="assembly_fragment"
	misc_feature	112919..163694	/note="assembly_fragment"
	misc_feature	clone_end:T7	
	misc_feature	vector_side:right"	
BASE COUNT	48221 a	33493 c	32231 g 48646 t 1103 others
ORIGIN			
Query Match	10.2%	Score 37.4;	DB 69; Length 163694;
Best Local Similarity	53.8%;	Pred. No. 0.71;	
Matches	77; Conservative	0; Mismatches	66; Indels 0; Gaps
Oy	112	acttagaacgtctgcgcgaagaaagtaagaagaaagtacatcctaccataaatata	171
Dn	139053	ACTTGAGACCTTAATCACTCTGGGCGCAAAATTGACTTTGCATCTAATGTCTTACTACTATAATA	138994
Oy	172	cactactgaatttcgltgatlggggaccttclgtttcatacaagcagtcactcatcgatctg	231
Dn	138993	GACTTCCTTAATCATCATATTATTTTCTTTTGTGAGACAGCTTGCCCTGTACACC	138934
Oy	232	atctgcatagcattatactatgc	254
Dn	138933	GGCTGCACTGCAGTGCAGTGATC	138911
RESULT 8			
AL450348/c	LOCUS	164055 bp	DNA HTG 01-FEB-2001
DEFINITION	Homo sapiens chromosome 6 clone RP11-149M20, *** SEQUENCING IN		
ACCESSION	AL450348 AC024499		
VERSION	AL450348.4 GI:12657205		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULITOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 164055)		
JOURNAL	Alameda,U.		
COMMENT	Direct Submission		
	Submitted (31-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone		
	requests: clonerequest@sanger.ac.uk		
	On Feb 2, 2001 this sequence version replaced gi:12192204.		
	Draft Sequence Produced by Whitehead Institute/MIT Center for		
	Genome Research, 320 Charles Street,		
	Cambridge, MA 02141, USA		
	http://www.seq.wi.mit.edu		
	----- Genome Center		
	Center: Sanger Centre		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: humquerry@sanger.ac.uk		
	Project Information		
	Center project name: BA149M20		
	----- Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Sequencing vector: M13; M77815; 49% of reads		
	Sequencing vector: plasmid; L08752; 50% of reads		
	Chemistry: Dye-terminator Big Dye; 99% of reads		
	Chemistry: Dye-primer-amersham; 0% of reads		
	Chemistry: Dye-primer Big Dye; 0% of reads		
	Consensus quality: I64022 bases at least Q40		

```

Consensus quality: 164045 bases at least Q30
Consensus quality: 164053 bases at least Q20
Insert size: 164055; sum-of-contigs
Insert size: 160390; 7.7% error; agarose-1p
Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
coverage: 9.44x in Q20 bases; agarose-1p
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
SOURCE
    Location/Qualifiers
        1..164055
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone_11b="RP11-149M20"
            /clone_11b="RP11-11.1"
            1..164055
            /note="assembly_fragment:01602
            clone_end:SP6
            vector_side:left
            clone_end:T7
            vector_side:right"

BASE COUNT      48936 a 33855 c 32515 g 48749 t

ORIGIN
Query Match      10.2%; Score 37.4; DB 81; Length 164055;
Best Local Similarity 53.8%; Pred. No. 0.71;
Matches 77; Conservative 0; Mismatches 66; Indels 0; Gaps 0

Oy 112 acttagaacgtctggcgtgaagtaaaagttaagagagctcctctcactacaataata 171
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139416 ACTTGGAACCTGTACTACTCTGGGGCCAAATTTGAGTTGACTCTTACTCTTACTCAATAATA 139357

Oy 172 cactactgaatttcgtgattgggggcctctgtttcactaagcagctgcagctactctgattcg 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139356 GACTTCTTAACTTCATCATCATTAATTTCTTTCTTTTGAGACAGCTGTGCCCCGTGCACCCG 139297

Oy 232 atctgcatagcattatctactgctc 234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139296 GGGTGGAGTGCAGTGCAGTGATC 139274

RESULT 9
AC020672/c LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-354C17 map 18, WORKING DRAFT
ACCESSION AC020672
VERSION AC020672.4 GI:10047733
KEYWORDS HMG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156838)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,L., Bouckhalter,B., Brown,A., Burrett,G., Castle,A.,
Choapel,Y., Colangelo,M., Collins,S.C., Collymore,A., Cooke,P.,
Deaertillano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Kleh,J.,
Landers,T., Lebeckzy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,
McDonald,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

```



```

VERSION      AC011451.6  GT:11863025
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE        1 (bases 1 to 180672)
REFERENCE    DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS      Direct Submission
TITLE        Unpublished
JOURNAL      2 (bases 1 to 180672)
REFERENCE    DOE Joint Genome Institute.
AUTHORS      Direct Submission
TITLE        Submitted (07-OCR-1999) Production Sequencing Facility, DOE Joint
JOURNAL      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3 (bases 1 to 180672)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE        Direct Submission
JOURNAL      Submitted (15-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
              On Dec 15, 2000 this sequence version replaced gi:10305127.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 100% of Sequence:
              Estimated Total Number of Errors is 0.
              90060-90070 : Tn insertion on the BAC removed
              148110 : 43bp PCR fragment sequence only
              STS Content:
              WI-14794 G32491
              SHGC-31653 G28548
              WI-30311 G22833.

FEATURES
  source
    1..180672
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="19"
    /clone="CTC-325H20"
    /clone="CTC-325H20"

BASE COUNT   51464 a 37679 c 38680 g 52849 t
ORIGIN
Query Match      9 8%; Score 35.6; DB 87; Length 180672;
Best Local Similarity 49.5%; Pred. No. 2.6;
Matches 92; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 119 actgtctgctgaagtaaaagtaagaagtcactctcactacaataatacactact 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 81812 ATTGATGGGATGTAATTAAGAGTATGGAAGGCAACAGGTGACAAATATATAACT 81753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 179 gaattcgatgattggggctctctgttcataagcaagtcgactactgtgactgtcga 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 81752 ATTTTATTAATTAAAGCATATGATTGAATTAAGCAACATTAAGTGTGACTCTCG 81693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 239 tagcatatactgctgctgactactgtcactagcactacaagctatctgactgca 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 81692 TATATAAGTCCAGGCTGTGATCTGTGCTTTACACACAGTACCATCTGTTTCTTACT 81633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 299 ttaggc 304
    |||||
DB 81632 GCAGCC 81627

```

```

ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE        1 (bases 1 to 171673)
REFERENCE    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL      Homo sapiens chromosome 2, clone RP11-231C6
AUTHORS      Unpublished
TITLE        2 (bases 1 to 171673)
REFERENCE    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Balwin,J., Barna,N., Bastien,V., Beda,F.,
              Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
              Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
              Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,
              Dodge,S., Domino,M., Doyle,M., Frettelra,P., Fitzhugh,W., Gage,D.,
              Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
              Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Iliev,I., Johnson,B., Jones,C., Kann,L., Kartas,A.,
              Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lechoczky,J.,
              Levine,R., Lileu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
              McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPheters,R.,
              Meldrim,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
              Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
              Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
              Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Testaye,S., Theodore,J., Titrrell,A., Travers,M., Triggillo,J.,
              Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
              Young,G., Zalnoun,J., Zimmer,A. and Zody,M.
              Direct Submission
              Submitted (07-JUL-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Aug 11, 2000 this sequence version replaced gi:18954163.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: 18033
              Center clone name: 231_C-6
              ----- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Assembly: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.9607.1
              Consensus quality: 16343 bases at least Q40
              Consensus quality: 167626 bases at least Q30
              Consensus quality: 169024 bases at least Q20
              Insert size: 182000; agarose-fp
              Insert size: 1695703; sum-of-contigs
              Quality coverage: 4.6 in Q20 bases; agarose-fp
              Quality coverage: 4.9 in Q20 bases; sum-of-contigs
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 22 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 5750: contig of 5750 bp in length
              * 5751 5850: gap of 100 bp
              * 5851 7251: contig of 1401 bp in length
              * 7252 7351: gap of 100 bp
              * 7352 8978: contig of 1627 bp in length
              * 8979 9078: gap of 100 bp
              * 9079 10926: contig of 1848 bp in length
              * 10927 11026: gap of 100 bp
              * 11027 12400: contig of 1374 bp in length

```

```

* 12401 12500: gap of 100 bp
* 12501 14741: contig of 2241 bp in length
* 14742 14841: gap of 100 bp
* 14842 17538: contig of 2697 bp in length
* 17539 17638: gap of 100 bp
* 17639 20714: contig of 3076 bp in length
* 20715 20814: gap of 100 bp
* 20815 23538: contig of 2724 bp in length
* 23539 23638: gap of 100 bp
* 23639 28315: contig of 4677 bp in length
* 28316 28415: gap of 100 bp
* 28416 36096: contig of 7681 bp in length
* 36097 36196: gap of 100 bp
* 36197 41940: contig of 5744 bp in length
* 41941 42040: gap of 100 bp
* 42041 48493: contig of 6453 bp in length
* 48494 48593: gap of 100 bp
* 48594 55611: contig of 7018 bp in length
* 55612 55711: gap of 100 bp
* 55712 65955: contig of 10244 bp in length
* 65956 66055: gap of 100 bp
* 66056 76390: contig of 10335 bp in length
* 76391 76490: gap of 100 bp
* 76491 87688: contig of 11198 bp in length
* 87689 87788: gap of 100 bp
* 87789 97788: contig of 10000 bp in length
* 97789 97888: gap of 100 bp
* 97889 111946: contig of 14058 bp in length
* 111947 112046: gap of 100 bp
* 112047 139941: contig of 27895 bp in length
* 139942 140041: gap of 100 bp
* 140042 159561: contig of 19520 bp in length
* 159562 159661: gap of 100 bp
* 159662 171673: contig of 12012 bp in length.

```

FEATURES

SOURCE

```

1. 171673
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-231C6"
/clone_id="RPCI-11 Human Male BAC"
1. 5750
/note="assembly-fragment
clone_end:SP6
vector_side:left"
misc_feature
5851..7251
/note="assembly-fragment"
7352..8978
/note="assembly-fragment"
9079..10926
/note="assembly-fragment"
11027..12400
/note="assembly-fragment"
12501..14741
/note="assembly-fragment"
14842..17538
/note="assembly-fragment"
17639..20714
/note="assembly-fragment"
20815..23538
/note="assembly-fragment"
23639..28315
/note="assembly-fragment"
28416..36096
/note="assembly-fragment"
36197..41940
/note="assembly-fragment"
42041..48493
/note="assembly-fragment"
48594..55611
/note="assembly-fragment"
55712..65955

```

```

/note="assembly-fragment"
66056..76390
/note="assembly-fragment"
76491..87688
/note="assembly-fragment"
87789..97788
/note="assembly-fragment"
97889..111946
/note="assembly-fragment"
112047..139941
/note="assembly-fragment"
140042..159561
/note="assembly-fragment"
159662..171673
/note="assembly-fragment
clone_end:T7
vector_side:right"
BASE COUNT 51122 a 34377 c 33665 g 50408 t 2101 others
ORIGIN

```

```

Query Match 9.6%; Score 35.2; DB 75; Length 171673;
Best Local Similarity 57.1%; Pred. No. 3.4;
Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```

```

QY 42 tgggtgactagatgtctcgtggcaaatcaatgaagcctcagcgtcaccgcgtt 101
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146605 TGGGTTCCACCGAGATTCCTCCCAAGACCTGGGTCCTCCCTCCCTGCTTT 146546
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 102 gtgcttgagactagactgtcgtcgtgaagtaaaagtaaaagta 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146545 CTTATTGCCACCATGACTCTGCGCTGTCCCAAGGGAAGGAAGTGA 146494

```

```

RESULT 14
AC010087 202029 bp DNA PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-241M13 from 2, complete sequence.
DEFINITION AC010087.3 GI:7630909
ACCESSION AC010087.3
VERSION AC010087.3
KEYWORDS hmg.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 202029)
AUTHORS Sulston,J.E. and Waterston,R.
JOURNAL Toward a complete human genome sequence
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
99063792
REFERENCE 2 (bases 1 to 202029)
AUTHORS Cordes,M., Walker,C., Drone,K. and Glaeser,E.
JOURNAL The sequence of Homo sapiens BAC clone RP11-241M13
unpublished
REFERENCE 3 (bases 1 to 202029)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (11-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 202029)
REFERENCE 4 (bases 1 to 202029)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 202029)
REFERENCE 5 (bases 1 to 202029)
AUTHORS Waterston,R.
JOURNAL Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 21, 2000 this sequence version replaced gi:5923729.
COMMENT ----- Genome Center

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0241M13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and co-workers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RPI1-617F9. Actual start of this clone is at base position 1 of RPI1-241M13; actual end is at base position 202029 of RPI1-241M13.

Location/Qualifiers

FEATURES

source

1. 202029
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RPI1-241M13"
/clone_id="RPI1-11"
49..173
/rpt_family="ERV"
repeat_region
174..234
/rpt_family="MaLR"
repeat_region
235..531
/rpt_family="Alu"
repeat_region
532..687
/rpt_family="MaLR"
repeat_region
688..990
/rpt_family="Alu"
repeat_region
991..1164
/rpt_family="MaLR"
repeat_region
1173..1532
/rpt_family="ERV"
repeat_region
1824..2092
/rpt_family="MaLR"
repeat_region
2201..2293
/rpt_family="MaLR"
repeat_region
2308..2492
/rpt_family="MaLR"
repeat_region
2796..3133
/rpt_family="MaLR"
repeat_region
3694..3801
/rpt_family="L1"

repeat_region 3802..3835
/rpt_family="MaLR"
repeat_region 3836..4181
/rpt_family="L1"
misc_feature 4318..5002
/note="similar to EST AL119394 (NID:g5925293)"
repeat_region 4319..4420
/rpt_family="Alu"
repeat_region 5077..5157
/rpt_family="MIR"
misc_feature 5739..5852
/note="similar to EST AM239056 (NID:g6571446) xb36a06.y1"
repeat_region 6305..6423
/rpt_family="MIR"
repeat_region 6476..6544
/rpt_family="MIR"
repeat_region 7257..7304
/rpt_family="MERL_type"
repeat_region 7305..7601
/rpt_family="Alu"
misc_feature 7371..7575
/note="similar to EST AA63642 (NID:g2617633) ae68c05.s1"
repeat_region 7602..7640
/rpt_family="MERL_type"
repeat_region 7653..7865
/rpt_family="MERL_type"
repeat_region 8242..8538
/rpt_family="Alu"
misc_feature 8256..8529
/note="similar to EST AA169801 (NID:g1748152) zp21d12.s1"
repeat_region 8539..8898
/rpt_family="L2"
repeat_region 9154..9333
/rpt_family="Alu"
repeat_region 9768..9867
/rpt_family="L2"
repeat_region 10126..10168
/rpt_family="MIR"
repeat_region 10188..10760
/rpt_family="L2"
repeat_region 10761..11050
/rpt_family="Alu"
repeat_region 11051..11644
/rpt_family="L2"
misc_feature 11618..11988
/note="similar to EST AM770140 (NID:g7702179) hk58c06.x1"
misc_feature 11621..12241
/note="similar to EST W41502 (NID:g1324639) mc50h09.r1"
misc_feature 11630..12189
/note="similar to EST A1822003 (NID:g5441082) o156e06.x5"
misc_feature 11634..12151
/note="similar to EST AM328089 (NID:g6798585) dr04b05.x1"
misc_feature 11637..12239
/note="similar to EST A1434076 (NID:g4293647) tl14g02.x1"
misc_feature 11637..12000
/note="similar to EST AA604021 (NID:g24444591) nc04f12.s1"
misc_feature 11637..11998
/note="similar to EST W45911 (NID:g1330618) mc81a10.r1"
misc_feature 11637..11914
/note="similar to EST AM372903 (NID:g6877557)"
misc_feature 11637..12095
/note="similar to EST AA721507 (NID:g2737642) nv54g06.r1"
misc_feature 11637..12244
/note="similar to EST AM245702 (NID:g6588695)"
misc_feature 11637..12064
/note="similar to EST AA056611 (NID:g1548951) z166d06.s1"
misc_feature 11637..11915
/note="similar to EST A1340462 (NID:g4077389) ta88c08.x1"
misc_feature 11637..11726
/note="similar to EST AM620042 (NID:g736226)"
misc_feature 11640..12041
/note="similar to EST AA542836 (NID:g2291316) n180e03.s1"
misc_feature 11640..12244

Wed May 9 10:01:10 2001

us-09-619-643-3.rge

Page 16

Search completed: May 8, 2001, 04:55:34
Job time: 6828 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:06:20 ; Search time 257.37 Seconds
(without alignments)
827.914 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatggcgcgtatcgtgtc.....ctaataccctacattcta 365

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_0401.*
1: /SIDS2/gcgdata/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.8	9.0	4630	20 V84139	Human prothrombina
2	31.6	8.7	1381	21 C33269	Arabidopsis thalia
3	31.6	8.7	78925	21 C89888	Human FN gene. Ho
4	31.4	8.6	1830121	17 T42063	Haemophilus influe
5	30.8	8.4	50000	21 A96363	Polymorphic repeat
6	30.6	8.4	1168	21 C34583	Arabidopsis thalia
7	30.6	8.4	1175	21 C34653	Arabidopsis thalia
8	30.6	8.4	1175	21 C46252	Arabidopsis thalia
9	30.4	8.3	3960	20 X83478	A.thaliana potassi
10	30.4	8.3	1664976	19 V21209	Methanococcus jann
11	29.8	8.2	1223	21 Z98349	A. thaliana gene 1

C 12	29.6	8.1	306	20 V90285	EST clone DJ188.
C 13	29.6	8.1	439	16 T20045	Human gene signatu
C 14	29.2	8.0	10391	20 X13070	Enterococcus faeca
C 15	28.8	7.9	855	20 X22240	Human secreted pro
C 16	28.8	7.9	1239	21 C45306	Arabidopsis thalia
C 17	28.8	7.9	1407	21 C35921	Arabidopsis thalia
C 18	28.8	7.9	3654	18 T72336	Lactobacillus bact
C 19	28.8	7.9	8738	18 T72337	Lactobacillus bact
C 20	28.6	7.8	1710	21 Z97103	Human secreted pro
C 21	28.6	7.8	1900	16 T05405	Schizosaccharomyce
C 22	28.6	7.8	99960	21 Z50905	Human TBC-1 partia
C 23	28.4	7.8	853	21 A59484	Nucleotide sequenc
C 24	28.4	7.8	6157	18 V74495	Staphylococcus aur
C 25	28	7.7	212	21 Z42604	Human 5' EST isola
C 26	28	7.7	333	20 Z18365	Microsatellite rep
C 27	28	7.7	333	21 A80263	Human ASTH1 locus
C 28	28	7.7	648	21 F13970	Aspergillus oryzae
C 29	28	7.7	762	20 X22120	Human secreted pro
C 30	28	7.7	3393	18 V74737	Staphylococcus aur
C 31	28	7.7	65921	21 Z89046	Human nbturin DNA.
C 32	28	7.7	72928	20 Z18355	Human ASTH1 5' ge
C 33	28	7.7	72928	21 Z18355	Human ASTH1 5' ge
C 34	27.8	7.6	571	20 A80263	Human ASTH1 5' ge
C 35	27.8	7.6	715	21 A01637	Rat neuronal limed
C 36	27.8	7.6	784	20 Z16328	Human colon cancer
C 37	27.6	7.6	1228	9 N80606	Human gene express
C 38	27.6	7.6	1228	12 Q10016	Lambda SM2M4 enco
C 39	27.6	7.6	1228	14 Q10016	Lambda SM2M4 enco
C 40	27.6	7.6	2436	22 A88924	Nucleotide sequenc
C 41	27.6	7.6	2692	22 A88923	Nucleotide sequenc
C 42	27.6	7.6	2788	22 A88922	Nucleotide sequenc
C 43	27.6	7.5	123025	20 X91990	Nucleotide sequenc
C 44	27.4	7.5	4189	10 N90570	Bovine repeat elem
C 45	27.2	7.5	1338	19 V14013	Xenopus cerberus c

ALIGNMENTS

RESULT 1	
ID V84139	standard; DNA: 4630 BP.
XX	
AC V84139:	
XX	
XX	15-MAR-1999 (first entry)
XX	
DE	Human prothrombinase hfg12 gene.
XX	
KW	Prothrombinase: hfg12; Fg12; human; immune coagulation; antibody;
KW	inhibitor; infection; graft rejection; glomerulonephritis; cancer;
KW	gastrointestinal disease; foetal loss; therapy; vaccine; ss.
XX	
OS	Homo sapiens.
XX	
PH	
FT	Key
FT	promoter
FT	
FT	5..1342
FT	location/Qualifiers
FT	protein_bind
FT	119..127
FT	/*tag= a
FT	/*tag= b
FT	/label= TCF1
FT	/note= "T cell factor 1 binding site"
FT	134..138
FT	protein_bind
FT	163..166
FT	/*tag= c
FT	/label= PEA3
FT	/note= "polyomavirus enhancer activator 3
FT	protein_bind
FT	181..189
FT	/label= GATTA1
FT	/*tag= e
FT	/label= NF_IL6

```

FT      /note="nuclear factor IL6 binding site"
FT      226..234
FT      /*tag= f
FT      /label= NF_IL6
FT      /note="nuclear factor IL6 binding site"
FT      301..307
FT      /*tag= g
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      376..380
FT      /*tag= h
FT      /label= NF_E1
FT      418..423
FT      /*tag= i
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      491..499
FT      /*tag= j
FT      /label= NF_IL6
FT      /note="nuclear factor IL6 binding site"
FT      506..510
FT      /*tag= k
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      553..558
FT      /*tag= l
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      564..569
FT      /*tag= m
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      572..577
FT      /*tag= n
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      637..645
FT      /*tag= o
FT      /label= NF_IL6
FT      /note="nuclear factor IL6 binding site"
FT      707..712
FT      /*tag= p
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      772..777
FT      /*tag= q
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      779..784
FT      /*tag= r
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      826..831
FT      /*tag= s
FT      /label= TCF1
FT      /note="T cell factor 1 binding site"
FT      844..849
FT      /*tag= t
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      1043..1049
FT      /*tag= u
FT      /label= NF_IL6
FT      /note="nuclear factor IL6 binding site"
FT      1083..1091
FT      /*tag= v
FT      /label= NF_IL6
FT      /note="nuclear factor IL6 binding site"
FT      1112..1117
FT      /*tag= w
FT      /label= bHLH
FT      /note="basic helix loop helix"
FT      1225..1233
FT      protein_bind

```

```

FT      /*tag= x
FT      /label= AABS
FT      1268..1274
FT      /*tag= y
FT      /label= API
FT      prim_transcript 1261
FT      /*tag= z
FT      /note="Transcription start site"
FT      TATA_signal
FT      1287..1293
FT      /*tag= aa
FT      prim_transcript 1261
FT      /*tag= ab
FT      /note="Transcription start site"
FT      CDS
FT      1343..1313
FT      /*tag= ac
FT      /note="contains an intron"
FT      exon
FT      1343..1955
FT      /*tag= ad
FT      /number= 1
FT      intron
FT      1956..2426
FT      /*tag= ae
FT      1956..2426
FT      /*tag= af
FT      /number= 2
FT      3'UTR
FT      3134..4187
FT      /*tag= ag

MO9851335-A1.
19-NOV-1998.
15-MAY-1998; 98MO-CA00475.
10-OCT-1997; 97US-0061684.
15-MAY-1997; 97US-0046537.
(LEVY/) LEVY G.
Levy G;
WPI: 1999-059687/05.
P-PSDB: W88235.

Modulating immune coagulation - by using Fg12 antibodies and
PT compounds, used to treat conditions including graft rejection and
foetal loss
PS Claim 8; Page 63-66; 105pp; English.
XX
XX This is the nucleotide sequence of the human prothrombinase gene
CC hfg12. To identify the gene, a human liver genomic DNA library
CC was subjected to PCR amplification using primers (see V84141-42)
CC specific to exon 2 of the mouse fg12 gene (see V84140). A 300 bp
CC PCR product was used to screen a PAC library. Clone 6360 was
CC isolated and subjected to restriction digestion to obtain clone
CC j14, which contained the hfg12 gene. The gene codes for human
CC prothrombinase fg12 (see W88235), a 70 kDa transmembrane serine
CC protease that has immune procoagulant activity. The invention
CC provides a method for inhibiting immune coagulation by inhibiting
CC the activity or expression of Fg12. The method can be used in vivo
CC to treat a condition which requires a reduction in immune
CC coagulation such as bacterial and viral infections, cancer,
CC glomerulonephritis, a number of gastrointestinal diseases,
CC allograft and xenograft rejection and foetal loss. An Fg12-specific
CC antibody, an Fg12 antisense oligonucleotide, or a substance that
CC affects prothrombinase activity of a Fg12 protein may be used to
CC treat a condition requiring a reduction in procoagulant activity.
CC A vaccine containing an Fg12 protein or peptide is used for
CC prevention of graft rejection or foetal loss (claimed).
XX
SQ Sequence 4630 BP; 1513 A; 820 C; 932 G; 1365 T; 0 other:

```


PR	28-MAY-1999;	9905-0136782;
PR	01-JUN-1999;	9905-0137122;
PR	03-JUN-1999;	9905-0137528;
PR	04-JUN-1999;	9905-0137502;
PR	07-JUN-1999;	9905-0137754;
PR	08-JUN-1999;	9905-0138094;
PR	10-JUN-1999;	9905-0138540;
PR	10-JUN-1999;	9905-0138847;
PR	14-JUN-1999;	9905-0139119;
PR	16-JUN-1999;	9905-0139452;
PR	16-JUN-1999;	9905-0139453;
PR	17-JUN-1999;	9905-0139454;
PR	18-JUN-1999;	9905-0139455;
PR	18-JUN-1999;	9905-0139456;
PR	18-JUN-1999;	9905-0139457;
PR	18-JUN-1999;	9905-0139750;
PR	18-JUN-1999;	9905-0139763;
PR	21-JUN-1999;	9905-0139817;
PR	22-JUN-1999;	9905-0139899;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140695;
PR	28-JUN-1999;	9905-0140823;
PR	29-JUN-1999;	9905-0140991;
PR	30-JUN-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141454;
PR	01-JUL-1999;	9905-0142152;
PR	02-JUL-1999;	9905-0142055;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0142920;
PR	12-JUL-1999;	9905-0142977;
PR	13-JUL-1999;	9905-0143542;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144005;
PR	16-JUL-1999;	9905-0144086;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144331;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144333;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0145086;
PR	21-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145192;
PR	22-JUL-1999;	9905-0145195;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145276;
PR	26-JUL-1999;	9905-0145913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146389;
PR	04-AUG-1999;	9905-0147204;

Db 11846 gatgagctcggtactcagagaagatgagatgtgtgaaggt 11887

RESULT 4
ID T42063 standard; DNA; 1830121 BP.

XX T42063;

XX 14-SEP-1999 (first entry)

XX Haemophilus influenzae complete genome sequence.

DE Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.

XX Haemophilus influenzae.

XX WO9633276-A1.

XX 24-OCT-1996.

XX 22-APR-1996; 96WO-US05320.

XX 07-JUN-1995; 95US-0487429.

XX 21-APR-1995; 95US-0426787.

XX 07-JUN-1995; 95US-0476102.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UXTD) UNITV JOHNS HOPKINS.

XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

XX WPI; 1996-485782/48.

XX Haemophilus influenzae Rd genome recorded on computer readable

XX medium - useful for identifying commercially important nucleic acid

XX fragments by homology searching

XX Claim 1; Page 77.2-77.1091; 1291pp; English.

XX This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.

XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 8.6%; Score 31.4; DB 17; Length 1830121;
Best Local Similarity 64.4%; Pred. No. 21;
Matches 47; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 cgaatggctcgaatgtgttcacagaccataaccctatatgtgtgactagattgct 60
DB 118454 cgaatggctcgaatgtgttcacagaccataaccctatatgtgtgactagattgct 118513

OY 61 ctggcgcaaatc 73
Db 118514 cggaaataatcc 118526

RESULT 5
A96363/c

ID A96363 standard; DNA; 50000 BP.

XX A96363;

XX 08-FEB-2001 (first entry)

XX Polymorphic repeat microsatellite sequences present in the CTLA4 locus.

XX Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CgRL; lupus;
KW insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
KW Hashimoto's disease; coeliac disease; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX 5722..5746

XX satellite

XX /tag= a

XX /note= "sara41/42 microsatellite repeat"

XX /tag= b

XX /note= "sara43/44 microsatellite repeat"

XX /tag= c

XX /note= "PW210/211 microsatellite repeat"

XX /tag= d

XX /note= "sara45/46 microsatellite repeat"

XX /tag= e

XX /note= "sara17/18 microsatellite repeat"

XX /tag= f

XX /note= "sara19/20 microsatellite repeat"

XX WO200056856-A2.

XX 28-SEP-2000.

XX 24-MAR-2000; 2000WO-US07938.

XX 25-MAR-1999; 99US-0126215.

XX (GEMV) GENETICS INST INC.

XX Ling V, Wu P, Gray GS;

XX WPI; 2000-628257/60.

XX Determining predisposition of humans to develop autoimmune disease

XX involves detecting polymorphic microsatellite repeat sequence within

XX human costimulatory receptor gene locus

XX Claim 2; Page 67-82; 160pp; English.

XX Two human bacterial artificial chromosome (BAC) clones that included
CC and flanked the human CTLA-4 locus were cloned and sequenced. The
CC sequence data was assembled into a continuous sequence that is presented
CC in A96363-68. A96363-64 comprise BAC clone 22700, and A96365-68 comprise
CC BAC clone 22608. The sequences contain polymorphic microsatellite repeat
CC (PMR) sequences. The specification describes a method for determining the
CC predisposition of a human subject to develop autoimmune disease. The
CC method comprises detecting a PMR sequence in the CD28, ICOS gene or
CC CTLA4 gene of the human costimulatory receptor gene locus (hCGR). PMR
CC sequences vary in length among individuals and can be amplified to
CC generate products that differ in size. These products can then be
CC detected by rapid and convenient high resolution processes. The method
CC is useful for determining the predisposition of insulin-dependent
CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 16-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.4%; Score 30.6; DB 21; Length 1168;
Best Local Similarity 45.6%; Pred. No. 1.6;
Matches 108; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 88 ggtacccgcgtgtgtcgtgtgtagactagaacgtctgctgaagcaaaagtaagga 147
Db 757 GGCATCCTTTGGCTTATGTGTGCATATTCCTGTGGGTGAACATCATATGTCGA 698

Qy 148 aggtactctcactacaataatacactactgaattcgtgattg99ggtcttctca 207
Db 697 ATCATACTTTGGCCAGGTTCATACCGTAATAATTTGAATGACCTCCATGGGACCTTG 638

Qy 208 taagcagtcgactactctgtatcgtatcgtatcgtatcgtatcgtatcgtatc 267
Db 637 GATCATGCTAGCTCTCTCGATTTCCTTCAACAAATCCAGACACAGTACTCTCTTC 578

Qy 268 tgcatagcactacaagctatgcatgcatgtagccctcctatggtcccaagaat 324
Db 577 TGAAGCACTGATTAAGTCTCTGAGCTGTCTAGTGTGCTTACTGTCTCAGCAGT 521

RESULT 7
C34663/c
ID C34663 standard; DNA; 1175 BP.
XX AC C34663;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7433.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX FE 25-FEB-2000; 2000EP-0301439.
XX XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135121.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139453.
PR 18-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 24-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.

PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 03-AUG-1999;	99US-0146389.
PR 04-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 05-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 06-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149176.
PR 20-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 23-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159295.
PR 13-OCT-1999;	99US-0159294.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159581.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161970.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match	8.48;	Score 30.6;	DB 21;	Length 1175;
Best Local Similarity	45.68;	Pred. No. 1.7;		
Matches 108; Conservative	0;	Mismatches 129;	Indels 0;	Gaps 0;

OY	88	ggaacccgagctgttgctgtgttgacttagagacgtcgtgtgagtaagaataaagga	147
Db	757	GGCAATCCCTTTGGCTGATTTGGTCCCATATTTCTGTGGGTGAACATCATAGTGGGA	696
OY	148	agtcactcttcactacaaatatacactactcgaatttcgtatgtgagggcttcgttca	207
Db	697	ATCATACCTTTTGGCCAGTTCATACCGTAAATTTGATGATACCTCCATGGACCTTGG	638
OY	208	taagcagtcgactactcctgactcgtactcgtacatactatgactcgtcgtactcgtac	267
Db	637	GATCATCTGACTCTCTTGGCATTTTCTTCAACAAAATCCAGACACAGATCACTCTCTC	578
OY	268	tgcactagactacaagaagcttatgctgactgacttagccctctatggtctccagaagt	324
Db	577	TGAAGCACTGATTAAGATTTCCCTGAGCTGTGTTCTAGTGCCCTTACGTGTTTACGAGGT	521

Result	ID	Accession	Gene	Gene ID	Gene Description
8	C46252/c	C46252 standard; DNA; 1175 bp.			
XX	AC	C46252;			
XX	DT	18-OCT-2000 (first entry)			
DE	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49459.			
XX	XX				
KW	KW	Hybridisation assay; genetic mapping; gene expression control			
KW	KW	protein identification; signal transduction pathway;			
KW	KW	metabolic pathway; promoter; termination sequence; ss.			
XX	OS	Arabidopsis thaliana.			
PN	PN	EP1033405-A2.			
XX	XX				
PD	PD	06-SEP-2000.			
XX	XX				
PE	PE	25-FEB-2000; 2000EP-0301439.			
XX	XX				
PR	PR	25-FEB-1999; 99US-0121825.			
PR	PR	05-MAR-1999; 99US-0123180.			
PR	PR	09-MAR-1999; 99US-0125348.			
PR	PR	23-MAR-1999; 99US-0125788.			
PR	PR	25-MAR-1999; 99US-0126264.			
PR	PR	29-MAR-1999; 99US-0126785.			
PR	PR	01-APR-1999; 99US-0127462.			
PR	PR	06-APR-1999; 99US-0128234.			
PR	PR	08-APR-1999; 99US-0128714.			
PR	PR	16-APR-1999; 99US-0128845.			
PR	PR	19-APR-1999; 99US-0130077.			
PR	PR	21-APR-1999; 99US-0130449.			
PR	PR	23-APR-1999; 99US-0130510.			
PR	PR	23-APR-1999; 99US-0130891.			
PR	PR	28-APR-1999; 99US-0131449.			
PR	PR	30-APR-1999; 99US-0132048.			
PR	PR	30-APR-1999; 99US-0132407.			

PR	04-MAY-1999	9905-013324.85
PR	05-MAY-1999	9905-013324.86
PR	06-MAY-1999	9905-013324.87
PR	07-MAY-1999	9905-013324.87
PR	08-MAY-1999	9905-013324.83
PR	11-MAY-1999	9905-013342.56
PR	14-MAY-1999	9905-013342.18
PR	14-MAY-1999	9905-013342.19
PR	14-MAY-1999	9905-013342.21
PR	14-MAY-1999	9905-013343.70
PR	18-MAY-1999	9905-013347.68
PR	19-MAY-1999	9905-013344.41
PR	20-MAY-1999	9905-013351.24
PR	21-MAY-1999	9905-013355.53
PR	22-MAY-1999	9905-013355.29
PR	25-MAY-1999	9905-013365.29
PR	27-MAY-1999	9905-013365.21
PR	28-MAY-1999	9905-013367.82
PR	01-JUN-1999	9905-013372.22
PR	03-JUN-1999	9905-013375.28
PR	04-JUN-1999	9905-013375.02
PR	07-JUN-1999	9905-013377.24
PR	08-JUN-1999	9905-013380.94
PR	10-JUN-1999	9905-013385.40
PR	10-JUN-1999	9905-013384.47
PR	14-JUN-1999	9905-013391.19
PR	15-JUN-1999	9905-013394.52
PR	16-JUN-1999	9905-013394.53
PR	17-JUN-1999	9905-013394.92
PR	18-JUN-1999	9905-013394.54
PR	18-JUN-1999	9905-013394.55
PR	18-JUN-1999	9905-013394.56
PR	18-JUN-1999	9905-013394.57
PR	18-JUN-1999	9905-013394.58
PR	18-JUN-1999	9905-013397.63
PR	21-JUN-1999	9905-013396.17
PR	22-JUN-1999	9905-013396.99
PR	23-JUN-1999	9905-014035.53
PR	23-JUN-1999	9905-014035.54
PR	24-JUN-1999	9905-014056.95
PR	28-JUN-1999	9905-014082.23
PR	29-JUN-1999	9905-014059.91
PR	30-JUN-1999	9905-014128.87
PR	01-JUL-1999	9905-014142.54
PR	01-JUL-1999	9905-014142.54
PR	02-JUL-1999	9905-014205.55
PR	06-JUL-1999	9905-014232.90
PR	08-JUL-1999	9905-014280.03
PR	16-JUL-1999	9905-014408.66
PR	19-JUL-1999	9905-014432.31
PR	19-JUL-1999	9905-014432.31
PR	19-JUL-1999	9905-014433.25
PR	19-JUL-1999	9905-014433.32
PR	19-JUL-1999	9905-014433.33
PR	19-JUL-1999	9905-014433.34
PR	19-JUL-1999	9905-014433.35
PR	20-JUL-1999	9905-014435.52
PR	20-JUL-1999	9905-014436.32
PR	20-JUL-1999	9905-014436.34
PR	21-JUL-1999	9905-014436.34
PR	21-JUL-1999	9905-014505.88
PR	21-JUL-1999	9905-014508.88

XX

XX

XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:05 ; Search time 132.33 Seconds
(without alignments)
481.599 Million cell updates/sec

Title: US-09-619-643-3

Perfect score: 365

Sequence: 1 cgatggcgcgattgtgttc.....ctaataccctacattcta 365

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2.6/prodata/2/lna/5A_COMB.seq: *
2: /cgn2.6/prodata/2/lna/5B_COMB.seq: *
3: /cgn2.6/prodata/2/lna/6A_COMB.seq: *
4: /cgn2.6/prodata/2/lna/6B_COMB.seq: *
5: /cgn2.6/prodata/2/lna/PCUTS_COMB.seq: *
6: /cgn2.6/prodata/2/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	28.4	7.8	853	4	US-09-247-373B-51
C 2	28	7.7	333	3	US-09-009-913-15
C 3	28	7.7	72928	3	US-09-009-913-1
C 4	27.6	7.6	1228	6	5187153-5
C 5	27.6	7.6	1228	6	5220013-5
C 6	27.6	7.6	1228	6	5220013-5
C 7	27.2	7.5	1411	4	US-08-878-474-2
C 8	27	7.4	138	3	US-08-075-520A-13
C 9	27	7.4	390	3	US-08-075-520A-16
C 10	27	7.4	504	3	US-08-075-520A-2
C 11	27	7.4	504	3	US-08-075-520A-3
C 12	27	7.4	534	3	US-08-075-520A-4
C 13	27	7.4	534	3	US-08-075-520A-5
C 14	27	7.4	558	3	US-08-075-520A-1
C 15	27	7.4	558	3	US-08-075-520A-25
C 16	27	7.4	588	3	US-08-075-520A-35
C 17	27	7.4	3220	2	US-08-225-488-1
C 18	27	7.4	6744	1	US-08-119-125A-2
C 19	27	7.4	35081	2	US-08-752-760A-1
C 20	26.8	7.3	5191	1	US-08-340-428B-1
C 21	26.8	7.3	5191	5	PCT-US93-07306-1
C 22	26.6	7.3	2025	3	US-09-160-483-2
C 23	26.6	7.3	2238	2	US-08-890-853-3
C 24	26.6	7.3	2238	2	US-09-099-125A-3
C 25	26.6	7.3	2238	2	US-09-099-125A-3
C 26	26.6	7.3	2273	2	US-09-197-360-1
C 27	26.4	7.2	2745	1	US-08-328-322-6

C 28	26.4	7.2	3500	1	US-08-328-322-4	Sequence 4, Appl
C 29	26.4	7.2	6042	5	US-08-261-822A-1	Sequence 1, Appl
C 30	26.4	7.2	6042	5	PCT-US95-07744A-1	Sequence 1, Appl
C 31	26.4	7.2	35524	3	US-08-823-137-1	Sequence 1, Appl
C 32	26.2	7.2	2201	2	US-08-974-565C-2	Sequence 2, Appl
C 33	26.2	7.2	2201	3	US-09-255-748-2	Sequence 2, Appl
C 34	26.2	7.2	2229	1	US-08-624-663A-1	Sequence 1, Appl
C 35	26.2	7.2	12047	2	US-09-022-461-1	Sequence 1, Appl
C 36	26	7.1	7493	1	US-08-212-133A-7	Sequence 7, Appl
C 37	26	7.1	7493	1	US-08-474-503-5	Sequence 5, Appl
C 38	26	7.1	7493	2	US-08-670-707A-5	Sequence 5, Appl
C 39	26	7.1	7493	3	US-09-037-601-5	Sequence 5, Appl
C 40	26	7.1	7493	5	PCT-US94-13200-5	Sequence 5, Appl
C 41	25.8	7.1	123	3	US-08-075-520A-12	Sequence 12, Appl
C 42	25.8	7.1	1236	2	US-08-955-138-15	Sequence 15, Appl
C 43	25.8	7.1	1877	2	US-08-955-138-11	Sequence 11, Appl
C 44	25.8	7.1	2332	1	US-08-252-492-1	Sequence 1, Appl
C 45	25.8	7.1	2332	2	US-08-727-126-1	Sequence 1, Appl

ALIGNMENTS

```
RESULT 1
US-09-247-373B-51/C
; Sequence 51, Application US/09247373B
; Patent No. 618954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247, 373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924, 747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 51
; LENGTH: 853
; TYPE: DNA
; ORGANISM: SOYBEAN
; US-09-247-373B-51

Query Match 7.8%; Score 28.4; DB 4; Length 853;
Best local Similarity 56.4%; Pred. No. 2;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 235 tgcatacatatctgctgctgctactctgatactgcatactgacataaagctatgca 294
    |||||
Db 362 TTCACACGGTTTCCTCTCTATCCCGCTATACCAATCATGCTGCTGATTA 303
    |||||

Qy 295 tgcattagccctctatgcttcacaaagatcag 328
    |||||
Db 302 AGCTTTCCTCAGCAATTAAGCCCAAAATCGAG 309
    |||||

RESULT 2
US-09-009-913-15/C
; Sequence 15, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axy's Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozilevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-15

Query Match      7.7%; Score 28; DB 3; Length 333;
Best Local Similarity 58.3%; Pred. No. 1.8;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 269 gcatagcactacaagaactatgtgcatgcatgaagccctctatggttcccaagaatcgag 328
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 GCCTCTACTACAGTCTCAGATATATCAAGTATGCTCTTATGATGAAGAAGTTTAAAG 75
QY 329 tgacatttaaaagtattcttaaac 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 TGTATCACAANAATCATCAACAC 51

RESULT 3
US-09-009-913-1/C
Sequence 1, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: Axy's Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESS: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-15
```

```
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match      7.7%; Score 28; DB 3; Length 72928;
Best Local Similarity 58.3%; Pred. No. 22;
Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 269 gcatagcactacaagaactatgtgcatgcatgaagccctctatggttcccaagaatcgag 328
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58566 GCCTCTACTACAGTCTCAGATATATCAAGTATGCTCTTATGATGAAGAAGTTTAAAG 58507
QY 329 tgacatttaaaagtattcttaaac 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58506 TGTATCACAANAATCATCAACAC 58483

RESULT 4
5187153-5/C
Patent No. 5187153
APPLICANT: CORDELL, BARBARA; SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER S
AMYLOID POLYPEPTIDE DERIVATIVES
NUMBER OF SEQUENCES: 33
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/502,273
FILING DATE: 29-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 361,912
FILING DATE: 06-JUN-1989
APPLICATION NUMBER: 359,911
FILING DATE: 12-MAY-1989
APPLICATION NUMBER: 87,002
FILING DATE: 18-AUG-1987
APPLICATION NUMBER: 8,810
FILING DATE: 30-JAN-1987
APPLICATION NUMBER: 948,376
FILING DATE: 31-DEC-1986
APPLICATION NUMBER: 932,193
FILING DATE: 17-NOV-1986
SEQ ID NO: 5
LENGTH: 1228
5187153-5

Query Match      7.6%; Score 27.6; DB 6; Length 1228;
Best Local Similarity 56.7%; Pred. No. 4.6;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 106 ttgttgactgaactgctgctgaagtaaaagttaaaggaaggtcaactctcaactaca 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 987 TTGAACCAAAAAGAAATGAAAGGTAAGGGAACCTGAAGAAGTAAGGCTTCTGCTGCTTTA 928
QY 166 aatacacactactgaatttcgtagatcgag 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 927 AATACATGCTATATATTTTGAAGAATGGGG 898

RESULT 5
5220013-5/C
Patent No. 5220013
APPLICANT: PONTE, PHYLLIS A.; CORDELL, BARBARA
```

Query Match	7.6%	Score 27.6	DB 6	Length 1228
Best Local Similarity	56.7%	Pred. No. 4.6		
Matches	51	Conservative	0	Mismatches 39
				Indels 0
				Gaps 0
QY 106	ttgtgacttgaactgcgtcgtgaagtaaaagttaaaggaaggtcattctactaca	165		
DB 967	TTGAACAAAAGAAATGAAAGGCTAAGGGAACCTGAAAGTAAAGCGCTTGCCCTTA	928		
QY 166	aatacactactgaattctcgtgattggg	195		
DB 927	AATCATGCTTAATTATTGAAAATGGGGG	898		

```

: RESULT      8
: US-08-075-520A-13
: Sequence 13, Application US/08075520A
: Patent No. 6020167
: GENERAL INFORMATION:
: APPLICANT: Thoma, Hans
: TITLE OF INVENTION: A Composition used as a
: TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
:

```


REGISTRATION NUMBER: 30099
REFERENCE/DOCKET NUMBER: 3757/MED1001US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 334-8991
TELEFAX: (612) 334-8994
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 bp
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-075-520A-2

Query Match 7.4%; Score 27; DB 3; Length 504;
Best Local Similarity 53.3%; Pred. No. 4.9;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Db 162 tacaatatatacactgaatttcgtagtgggggcttcgtttcataagcagtcgacta 221
241 TCCAGAGATCTAGTACCAATTATGTTAACTAACAATGGGTTTAAAGATCAGGCAACTA 300

Qy 222 ctctgactgactgcataagcatatactgctgcgtctactctgactc 268
301 TTGTGCTTCATATATCTGCTTACTTTTGAAGAGACTGTACT 347

Db 301 TTGTGCTTCATATATCTGCTTACTTTTGAAGAGACTGTACT 347

RESULT 11
US-08-075-520A-3
Sequence 3, Application US/08075520A
Patent No. 6020167
GENERAL INFORMATION:
APPLICANT: Thoma, Hans
TITLE OF INVENTION: A Composition Used as a
TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas E. Popovich, Thomas Popovich
STREET: 80 South 8th Street
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55402-2111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
COMPUTER: IBM Compatible Compaq Prolinea 4/66
OPERATING SYSTEM: MS-DOS Version 5
SOFTWARE: Microsoft Word for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/075,520A
FILING DATE: January 31, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124775
FILING DATE: December 19, 1990
APPLICATION NUMBER: PCT/EP91/02460
FILING DATE: December 19, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Thomas E. Popovich
REGISTRATION NUMBER: 30099
REFERENCE/DOCKET NUMBER: 3757/MED1001US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 334-8991
TELEFAX: (612) 334-8994
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 bp
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-075-520A-3

Query Match 7.4%; Score 27; DB 3; Length 504;
Best Local Similarity 53.3%; Pred. No. 4.9;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Db 162 tacaatatatacactgaatttcgtagtgggggcttcgtttcataagcagtcgacta 221
241 TCCAGAGATCTAGTACCAATTATGTTAACTAACAATGGGTTTAAAGATCAGGCAACTA 300

Qy 222 ctctgactgactgcataagcatatactgctgcgtctactctgactc 268
301 TTGTGCTTCATATATCTGCTTACTTTTGAAGAGACTGTACT 347

Db 301 TTGTGCTTCATATATCTGCTTACTTTTGAAGAGACTGTACT 347

RESULT 12
US-08-075-520A-4
Sequence 4, Application US/08075520A
Patent No. 6020167
GENERAL INFORMATION:
APPLICANT: Thoma, Hans
TITLE OF INVENTION: A Composition Used as a
TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas E. Popovich, Thomas Popovich
STREET: 80 South 8th Street
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55402-2111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
COMPUTER: IBM Compatible Compaq Prolinea 4/66
OPERATING SYSTEM: MS-DOS Version 5
SOFTWARE: Microsoft Word for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/075,520A
FILING DATE: January 31, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124775
FILING DATE: December 19, 1990
APPLICATION NUMBER: PCT/EP91/02460
FILING DATE: December 19, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Thomas E. Popovich
REGISTRATION NUMBER: 30099
REFERENCE/DOCKET NUMBER: 3757/MED1001US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 334-8991
TELEFAX: (612) 334-8994
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 534 bp
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-075-520A-4

Query Match 7.4%; Score 27; DB 3; Length 534;
Best Local Similarity 53.3%; Pred. No. 5;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 162 tacaatatatacactgaatttcgtagtgggggcttcgtttcataagcagtcgacta 221
271 TCCAGAGATCTAGTACCAATTATGTTAACTAACAATGGGTTTAAAGATCAGGCAACTA 330

Qy 222 ctctgactgactgcataagcatatactgctgcgtctactctgactc 268
301 TTGTGCTTCATATATCTGCTTACTTTTGAAGAGACTGTACT 347

```

Db      331  TTGGGCTTCAATATCTTGCCTTACTTTTGGAGAGAGACTGTA 377

RESULT  13
US-08-075-520A-5
: Sequence 5, Application US/08075520A
: Patent No. 6020167
: GENERAL INFORMATION:
: APPLICANT: Thoma, Hans
: TITLE OF INVENTION: A Composition Used as a
: TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Thomas E. Popovich, Thomas Popovich
: STREET: 80 South 8th Street
: CITY: Minneapolis
: STATE: Minnesota
: COUNTRY: USA
: ZIP: 55402-2111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
: MEDIUM TYPE: Storage
: COMPUTER: IBM Compatible Compaq Prolinea 4/66
: OPERATING SYSTEM: MS-DOS Version 5
: SOFTWARE: Microsoft Word for Windows
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/075, 520A
: FILING DATE: January 31, 1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 90124775
: FILING DATE: December 19, 1990
: APPLICATION NUMBER: PCT/EP91/02460
: FILING DATE: December 19, 1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Thomas E. Popovich
: REGISTRATION NUMBER: 30099
: REFERENCE/DOCKET NUMBER: 3757/MED1001US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (612) 334-8991
: TELEFAX: (612) 334-8994
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 534 bp
: TYPE: Nucleic Acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: US-08-075-520A-5

Query Match      7.44; Score 27; DB 3; Length 534;
Best Local Similarity 53.3%; Pred. No. 5;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY      162  tacaatatcactactgaatttcgtgtgttggtggggtctctgtttcattcaagcagtcgacta 221
Db      271  TCCAGAGATCTAGTAGTCATTAATTATGTTATCTAACAATGGGTTTAAAGATGAGCACTA 330
QY      222  ctctgattcgtatctgcacatagcatatactggtcggtactctgact 268
Db      331  TTGGGCTTCAATATCTTGCCTTACTTTTGGAGAGAGACTGTA 377

RESULT  14
US-08-075-520A-1
: Sequence 1, Application US/08075520A
: Patent No. 6020167
: GENERAL INFORMATION:
: APPLICANT: Thoma, Hans
: TITLE OF INVENTION: A Composition Used as a
: TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
: NUMBER OF SEQUENCES: 40

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Thomas E. Popovich, Thomas Popovich
ADDRESSEE: 6 Associates
STREET: 80 South 8th Street
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55402-2111

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
MEDIUM TYPE: Storage
COMPUTER: IBM Compatible Compaq Prolinea 4/66
OPERATING SYSTEM: MS-DOS Version 5
SOFTWARE: Microsoft Word for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/075_520A
FILING DATE: January 31, 1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124775
FILING DATE: December 19, 1990
APPLICATION NUMBER: PCT/EP91/02460
FILING DATE: December 19, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Thomas E. Popovich
REGISTRATION NUMBER: 30099
REFERENCE/DOCKET NUMBER: 3757/MED1001US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 334-8991
TELEFAX: (612) 334-8994
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 bp
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-075-520A-1

Query Match 7.4%; Score 27; DB 3; Length 558;
Best Local Similarity 53.3%; Pred. No. 5.1;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 162 tacaataatacactgaattcgtgattcgaggcctcgttctcataagcagtcgacta 221
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 tccaaagatctcagatgctgcattatgcttaataactaaacatggctttaaacatcagcgaacta 300
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 222 ctctgactcgtacatcgatcatatatactatcgtgcgctactctgactc 268
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 ttgtgctttcaatattatcttgccttacttttggaagacagactgact 347
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
US-08-075-520A-25
; Sequence 25, Application US/08075520A
; Patent No. 6020167
; GENERAL INFORMATION:
; APPLICANT: Thoma, Hans
; TITLE OF INVENTION: A Composition Used as a
; TITLE OF INVENTION: Therapeutic Agent Against Chronic Viral Hepatic Diseases
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas E. Popovich, Thomas Popovich
; ADDRESSEE: 6 Associates
; STREET: 80 South 8th Street
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: USA
; ZIP: 55402-2111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible Compaq Prolinea 4/66

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:26 ; Search time 3095.41 Seconds
(without alignments)
1030.130 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatggcgtcgtgtgttc.....ctatactccctacattcta 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: em_gss_fun:*
 169: em_gss_hum1:*
 170: em_gss_hum2:*
 171: em_gss_hum3:*
 172: em_gss_hum4:*
 173: em_gss_hum5:*
 174: em_gss_hum6:*
 175: em_gss_hum7:*
 176: em_gss_hum8:*
 177: em_gss_hum9:*
 178: em_gss_inv1:*
 179: em_gss_inv2:*
 180: em_gss_inv3:*
 181: em_gss_other:*
 182: em_gss_pnl1:*
 183: em_gss_pnl2:*
 184: em_gss_pro1:*
 185: em_gss_pro2:*
 186: em_gss_pro3:*
 187: em_gss_pro4:*
 188: em_gss_pro5:*
 189: em_gss_pro6:*
 190: em_gss_pro7:*
 191: em_gss_pro8:*
 192: em_gss_pro9:*
 193: em_gss_pro10:*
 194: em_gss_pro11:*
 195: em_gss_pro12:*
 196: em_gss_pro13:*
 197: em_gss_pro14:*
 198: em_gss_pro15:*
 199: em_gss_pro16:*
 200: em_gss_pro17:*
 201: em_gss_pro18:*
 202: em_gss_pro19:*
 203: em_gss_pro20:*
 204: em_gss_pro21:*
 205: em_gss_pro22:*
 206: em_gss_pro23:*
 207: em_gss_pro24:*
 208: em_gss_pro25:*
 209: em_gss_pro26:*
 210: em_gss_pro27:*
 211: em_gss_pro28:*
 212: em_gss_pro29:*
 213: em_gss_pro30:*
 214: em_gss_pro31:*
 215: em_gss_pro32:*
 216: em_gss_pro33:*
 217: em_gss_pro34:*
 218: em_gss_pro35:*
 219: em_gss_pro36:*
 220: em_gss_pro37:*
 221: em_gss_pro38:*
 222: em_gss_pro39:*
 223: em_gss_pro40:*
 224: em_gss_pro41:*
 225: em_gss_pro42:*
 226: em_gss_pro43:*
 227: em_gss_pro44:*
 228: em_gss_pro45:*
 229: em_gss_pro46:*
 230: em_gss_pro47:*
 231: em_gss_pro48:*
 232: em_gss_pro49:*
 233: em_gss_pro50:*
 234: em_gss_pro51:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	41.2	11.3	216	AA231922	AA231922 CD0520.F
2	41.2	11.3	485	BE439089	BE439089 CD0520.MH
3	39.8	10.9	505	BE024097	BE024097 sm96f09.y
4	39.8	10.9	603	BE329865	BE329865 so70e11.y
5	38	10.4	354	BE439163	BE439163 CD01241.W
6	37.8	10.4	341	AO806290	AO806290 HS_3168.A
7	37.2	10.2	939	BF102078	BF102078 601752759
8	35.2	9.6	379	AZ275759	AZ275759 RPCI-23-1
9	34.4	9.4	976	AO900980	AO900980 HS_3233.B
10	33.6	9.2	595	AZ606891	AZ606891 IM0429K01
11	33.2	9.1	426	H95649	H95649 yv17c05.41
12	33.2	9.1	1101	CNS00818	AL057210 Drosoph11
13	33	9.0	391	AQ368869	AQ368869 HS_5021.B
14	33	9.0	406	AZ618773	AZ618773 IM0450H04
15	33	9.0	869	AZ687120	AZ687120 ENTGAB85FR
16	33	9.0	223	AZ539574	AZ539574 ENTGFI6FR
17	32.8	9.0	440	AO594432	AO594432 HS_2105.B
18	32.8	9.0	856	BE787147	BE787147 601476806

```

c 19 32.6 8.9 520 165 BE329768 BE329768 soc69e02.y
c 20 32.4 8.9 461 115 AM462113 BP230008A
c 21 32.4 8.9 540 168 BF728175 BF728175 1000057A0
c 22 32.4 8.9 670 204 AQ254315 AQ254315 Cp60761B
c 23 32.2 8.8 300 7 AA407249 AA407249 EST02266
c 24 32.2 8.8 802 229 CNS01241 AL101100 Drosophila
c 25 31.8 8.7 455 209 AO634564 AO634564 RPCI-11-4
c 26 31.8 8.7 590 3 AA153223 AA153223 mt17b04.t
c 27 31.6 8.7 238 167 BE439045 BE439045 CD0520.WH
c 28 31.6 8.7 519 206 AQ450981 AQ450981 HS_5152_A
c 29 31.6 8.7 584 214 AQ082356 AQ082356 RPCI-23-3
c 30 31.6 8.7 852 225 AZ669068 AZ669068 ENTIN1TR
c 31 31.6 8.7 870 225 AZ672956 AZ672956 ENTKCS8TF
c 32 31.6 8.7 979 225 AZ671363 AZ671363 ENRT96TF
c 33 31.6 8.7 1101 229 CNS00175 AL078740 Drosophila
c 34 31.4 8.6 345 115 AM430197 AM430197 69349 MAR
c 35 31.4 8.6 461 203 AQ026722 AQ026722 HS_2023_B
c 36 31.4 8.6 481 211 AQ0793825 AQ0793825 HS_5252_B
c 37 31.4 8.6 886 231 CNS04K6N AL294440 Tetradon
c 38 31.4 8.6 934 225 AZ671921 AZ671921 ENTHL85TR
c 39 31.4 8.6 967 145 BE214585 BE214585 601845893
c 40 31.4 8.6 987 210 AQ0744251 AQ0744251 HS_5508_A
c 41 31.4 8.6 1003 225 AZ671243 AZ671243 ENTLAS8TR
c 42 31.4 8.6 1101 229 CNS00016 AL058202 Drosophila
c 43 31.2 8.5 400 203 AQ243970 AQ243970 HS_2058_B
c 44 31.2 8.5 438 150 BE600693 BE600693 265444 MA
c 45 31.2 8.5 474 214 AO997942 AO997942 RPCI-23-3

```

ALIGNMENTS

```

RESULT 1
LOCUS AA231922/c 216 bp mRNA 15-SEP-2000
DEFINITION CD0520.F cDNA from oat Avena sativa cDNA clone CD0520, mRNA
ACCESSION AA231922
VERSION AA231922
KEYWORDS EST.
SOURCE oat.
ORGANISM Avena sativa
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Poae; Avena.
1 (bases 1 to 216)
AUTHORS Vandeynze,A.E., Sorrells,M.E., Park,W.D., Ayres,N.M., Fu,H.,
Cartinhou,S.W., Paul,E. and McCouch,S.R.
TITLE Anchor Probes for Comparative Mapping of Grass Genera
JOURNAL Theor. Appl. Genet. 97, 356-369 (1998)
COMMENT Contact: McCouch SR
Dept Plant Breeding
Cornell University
Ithaca, NY 14853-1901, USA
Tel: 607 255 0420
Fax: 607 255 6683
Email: srm4@cornell.edu
CDNA from oat (Avena sativa); forward sequence of RFLP probe
CD0520. Sequence determined by Nicola M. Ayres. For mapping
information, additional citations and other related information
concerning this probe, please refer to the Ricegenes database at
http://ars-genome.cornell.edu/cgi/wdbin/Webace?db=ricegenes&
class=MarkerObject-CD0520.
FEATURES
source
location/Qualifiers
1..216
/organism="Avena sativa"
/cultivar="Brooks"
/db_xref="Ricegenes:CD0520"
/db_xref="taxon:4498"
/clone="CD0520"
/clone_lib="cDNA from oat"
/tissue_type="etiolated leaf"
/note="Vector: Uni-ZAP XR/Pluescript; Site_1: EcoRI;
Site_2: XhoI; A Uni-ZAP XR cDNA library was constructed

```

```

BASE COUNT 62 a 59 c 36 g 59 t
ORIGIN
Query Match 11.3%; Score 41.2; DB 4; Length 216;
Best Local Similarity 70.5%; Pred. No. 0.023;
Matches 55; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```

```

Qy 2 gatggatcgattgtgtccagagccataaccatattggtgactagattgtc 61
||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 GATGGCAGCTGCTGCTGCGAGCCGCTGACATGGGTGGGCTTGGGCTCCTC 141
Qy 62 tgggcaaatcagaatga 79
Db 140 TCCACAAACCAAGTGA 123

```

```

RESULT 2
LOCUS BE439089/c 485 bp mRNA 24-JUL-2000
DEFINITION CD0520.WHE2F0032F ITTC CDO Oat cDNA library Avena sativa cDNA clone
CD0520.WHE2, mRNA sequence.
ACCESSION BE439089
VERSION BE439089
KEYWORDS EST.
SOURCE oat.
ORGANISM Avena sativa
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Poae; Avena.
1 (bases 1 to 485)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Izzo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pechioni,N., Quailset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticaceae EST Cooperative (ITTC): Production of
Expressed Sequence Tags for Species of the Triticaceae
JOURNAL Unpublished (2000)
COMMENT Contact: Sorrells M
Dept. of Plant Breeding & Biometry, Cornell University
Ithaca, NY 14853, USA
Tel: 607 255 1665
Fax: 607 255 6683
Email: mes12@cornell.edu
International Triticaceae EST Cooperative (ITTC)
http://wheat.pw.usda.gov/genome.
FEATURES
source
location/Qualifiers
1..485
/organism="Avena sativa"
/db_xref="taxon:4498"
/clone="CD0520.WHE2"
/clone_lib="ITTC CDO Oat cDNA library"
/tissue_type="leaf"
/note="Vector: PluescriptSK(-); This probe has been used
for mapping studies for species of the Triticaceae and is
available from the GrainGenes Probe Repository (
http://wheat.pw.usda.gov )"
BASE COUNT 109 a 127 c 108 g 141 t
ORIGIN
Query Match 11.3%; Score 41.2; DB 167; Length 485;
Best Local Similarity 70.5%; Pred. No. 0.029;
Matches 55; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```

Oy	2	gatggcgcgatgttggtccagacgaccataactaaccttgggtgacgaagtcttc 61 +
Db	224	GATGGCACTGTGGTTCGAGACCGCTGAAGCCCTACATTGGGTGGGTTCTC 165 +
Oy	62	tgggcaaatccaagtga 79 +
Db	164	TCCCAAAACCAAGTGA 147
RESULT	3	
BE024097/c		
LOCUS		
DEFINITION		BE024097 505 bp mRNA EST 06-JUN-2000
ACCSSION		sm86f09.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
VERSION		Gm-cl015-7938 5' similar to TR:080443 O80443 F16M14.11 PROTEIN. ;,
KEYWORDS		mRNA sequence.
SOURCE		BE024097
ORGANISM		BE024097.1 GI:8286538
REFERENCE		EST. soybean. glycine max Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine. 1 (bases 1 to 505) Shoemaker,R., Keim,P., Vodkin,L., Erpelnding,J., Coryell,V., Khanna, 'A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Stepienoe,M., Thelsting,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., Mccann ,R., Waterston,R. and Wilson,R. Public Soybean EST project Unpublished (1999)
JOURNAL		Contact: Shoemaker R./Publie Soybean EST project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Seq primer: -40RP from Glbco High quality sequence sTop: 393. Location/Qualifiers 1..505 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-7938" /clone_1lb="Gm-cl015" /tissue_type="Mature flowers, field grown plants" /lab_host="XLI0-Gold" /note="Vector: pBluescript II XR; Site1: EcoRI; Site2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapter were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XLI0-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelnding."
BASE COUNT	147 a	87 c 122 g 149 t
ORIGIN		
Query Match		10.9%; Score 39.8; DB 161; Length 505;

Best Local Similarity 56.5% ; Pred. No. 0.08 ; Mismatches 74; Conservative 0; Gaps 0;

OY 158 tcactacaataacatcacgcgaatttcgtagtatggsgcgcttctgtttccataagcagtcg 217
 | | | | |
Db 219 TAACACCACAATTTCCTGCGTCCGCAGCCTGTGAATGGATGCCATCATCAAGAACAAGCATTTCA 160

Oy 218 actactcgcagtcatgcgatcatactatcacttgctcgacctatctga.ctcyctaagaac 277
 | | | | |
Db 159 TCCACTCATCTCTCTCTTGCGAGTGCATTAATAAAGATCAACAACATTGTGACATCACACTTCCCT 100

OY 278 tacaaggctta 288
 | | | | |
Db 99 TGCAATAGCTTA 89

RESULT 4
BE329865/c

LOCUS
DEFINITION
scf7Oell.yl.Gm-cl040 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
gm-cl040-1029 5' similar to FR:080443 OR0443 FL6M14.II PROTEIN.;,
mrna sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
soybean.
Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 603)

REFERENCE
AUTHORS
Shoemaker,R., Kelm,P., Vodkin,L., Eredelding,J., Coryell,V., Khanna
. A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,N., Bowers
. Y., Person,B., Swaller,T., Gibbons.M., Page,D., Harvey.N., Schurk
. R., Ritter,E., Korn,S., Shn.T., Jackson,Y., Cardenas.M., McCann
. R., Waterston.R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R./public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108, USA
Tel.: 314 286 1800
Fax: 314 286 1810
Email: estwatson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 546 Std Error: 0.00
High quality sequence stop: 472.
Location/Qualifiers

1. 603
/organism="Glycine max"
/db_xref=taxon:3847"
/clone=GENOME SYSTEMS CLONE ID: Gm-cl040-1029"
/contig_lib="Gm-cl040"
/tisue_type="Hypocotyl and Plumule, germinating seeds"
/lab_host="DH10B"
/note=Vector: pTR73pac (Pharmacia); site_1: EcoRI;
site_2: NotI. This cdna library was constructed from mrRNA
isolated from hypocoty1 and plumule tissues of seeds
Complimentary DNA was synthesized from mRNA using a prime
consisting of a poly(dT) sequence with a NotI restriction
site. EcoRI adapters were ligated to the blunt-ended cdN
fragments followed by digestions with EcoRI and NotI. Th
cdNA fragments were directionally cloned into the
EcoRI-NotI restriction site of the pTr73-Pac vector. The
ligated cdNA fragments were transformed into DH10B host
cells (GldCo BR). This library was constructed by Dr.

DEFINITION	601752759F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3980174 5 mRNA sequence.					
ACCESSION	BF102078					
VERSION	BF102078.1 GI:10884604					
KEYWORDS	EST.					
ORGANISM	house mouse. Mus musculus					
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 939)					
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .					
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM9175 row: f column: 15 High quality sequence stop: 537.					
FEATURES	location/Qualifiers					
SOURCE	1..939					
	/organism="Mus musculus"					
	/strain="FVB/N"					
	/db_xref="taxon:10090"					
	/clone="IMAGE:3980174"					
	/clone_1lb="NCI_CGAP_Mam1"					
	/tissue_type="tumor, biopsy sample"					
	/dev_stage="3 months, virgin"					
	/lab_host="DH10B"					
	/note="Organ: mammary; Vector: PCMY-SPORE; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"					
BASE COUNT	247 a 173 c 261 g 258 t					
ORIGIN						
Query Match	10.2% Score 37.2; DB 144; Length 939;					
Best Local Similarity	56.6%; Pred. No. 0.63;					
Matches 69; Conservative	0; P-matches 53; Indels 0; Gaps 0					
OY	82 ctcttcaggtcacgcgctgtgtcgctgtgactagaactgtcgtgctaagtaaaagt 141					
DB	503 CGCTACTGTCCCTTTCTGTCTGTATTTTGAAAACCTTTAGTGGCTGAAGAAAAATT 444					
OY	142 aaagggaagttcaattctcatcaaatatcacactactgaatttcgtgatggggacttc 201					
DB	443 AACCGAATAAATAATTCCAACAACGTAACACAGCTTATATCTTCATTTACGATTCTC 384					
OY	202 gt 203					
Db	383 AT 382					
RESULT	8					
LOCUS	A2275759/c 379 bp DNA GSS 26-JUL-2000					
DEFINITION	RPI1-23-107L5.TJ RPI1-23 Mus musculus genomic clone RPI1-23-107L5, DNA sequence.					
ACCESSION	A2275759					
VERSION	A2275759.1 GI:9494661					
KEYWORDS	GSS.					
SOURCE	house mouse. Mus musculus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 379)					

TITLE	JOURNAL	COMMENT
Unpublished (1999)	Other:GSSS: RPCI-23-107L5.TV	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)		
Seg primer: SP6	Plate: 107 row: L column: 5	Class: BAC ends.
Location/Qualifiers		
1. .379	/organism="Mus musculus"	/strain="C57BL/6J"
	/db_xref="taxon:10090"	/clone="RPCI-23-107L5"
	/clone_1Id="RPCI-23"	/sex="Female"
	/lab_host="DH10B"	/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT	90 a 63 c 70 g 156 t	
ORIGIN		
Query Match	9.6%; Score 35.2; DB 218; Length 379;	
Best Local Similarity	55.8%; Pred. NO. 2.1;	
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;		
QY 98	tgttgccttgtagacttagacgtctgcgtgtaagtaaaagttaaaggaagtcacttc 157	
Db 315	TGGCTTTCTCTGCGACAAAGGACAAAGCGCTGAAAAGAAATTAAGGAAATACACCCC 256	
QY 158	tcaactacaatacactacactgaatttcgtgtagtgggggctctgttccataagcagtcg 217	
Db 255	TCAGAAATAATATTAATAAACTGCTAGCGGCTGTAGTGCACACCTTAATCCACGACTTGG 196	
RESULT 9		
LOCUS	AQ900980	
DEFINITION	HS_3233_B2_A03_TTC CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3233 Col=6 Row=B, DNA sequence.	
ACCESSION	AQ900980	
KEYWORDS	AQ900980.1 GI:6357170	
SOURCE	GSS.	
ORGANISM	human.	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 976)	
AUTHORS	Mahtaris,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

TITLE	THE WASHU-MERCK EST PROJECT
JOURNAL	UNPUBLISHED (1995)

COMMENT	FEATURES	ORIGIN
Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu High quality sequence stops: 266 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 887 Std Error: 0.00 Seq primer: ml3 -40 forward.	source	
	location/Qualifiers	
	1..426	
	/organism="Homo sapiens"	
	/db_xref="GDB:3792117"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:242984"	
	/clone_lib="Soares fetal liver spleen INPLS"	
	/sex="male"	
	/dev_stage="20 week post conception fetus"	
	/lab_host="DH10B (ampicillin resistant)"	
	/note="Organ: Liver and spleen. Vector: pTY73D (Pharmacia) with a modified polylinker. Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTGCAGAGATTAATTAAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTY73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	136 a 76 c 75 g 137 t	2 others
ORIGIN		
Query Match	9.1%; Score 33.2; DB 154; Length 426;	
Best Local Similarity	50.6%; Pred. No. 9.1;	
Matches	80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;	
OY 134	taaaagttaaaggaaggtcactcttcactacaatacactactgaatttcgtagttg 193	
Db 54	TTAAATTAAATCTCAAAATCTTTGTACTAGTAGACAAATGATGCTTTCTTTTAA 113	
OY 134	gggcttcggttcataaagcagtcgactctctgcatactgatactagcatatatactg 253	
Db 114	ATTCTATCTGTTTAAATAACTGTCAACATCATAGTATCTCTTGAGTAAATTGTGAGTGGG 173	
OY 254	gggctactctgcatcgcataagcaaacagactatgt 291	
Db 174	CAGCATCTGTACTTTTTCAGAAAGGAGAAAGATAGATATTT 211	
RESULT 12		
CNS00B18		
LOCUS	CNS00B18 1101 bp DNA GSS 04-JUN-1999	
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #	
	BACR32M20 of RP11-98 library from Drosophila melanogaster (fruit	
	fly), genomic survey sequence.	
ACCESSION	AL057210	
VERSION	AL057210.1 GI:4937676	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 1101)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)	
COMMENT	Determination of this BAC-end sequence was carried out as part of a	

FEATURES	source	Location/Qualifiers
BASE COUNT	352 a 192 c 150 g	241 t 166 others
ORIGIN	<pre> 1..1101 /db.xref="taxon:7227" /clone_lib="RPC1-98" /clone="BACR23M20" /note="End : TET3" </pre>	
Query Match	9.1%, Score 33.2, DB 229, Length 1101;	
Best Local Similarity	41.3%, Pred. No. 12;	
Matches	83; Conservative	17; Mismatches 101; Indels 0; Gaps 0;
QY	<pre> 160 actcaaaatatactactcgaatcgcgatgctggggccttcggtccataagcagtcgac 219 DB 458 ACTACAAATATTCACAAATTTACAAATCKRMAAAATTKCTCTTCATTTAATTAATWCTT 517 QY 220 tactctgacatcgtatcgtacagcaltatactgctgcgtactcgtatcgtacagcacta 279 DB 518 TACTCTCGTCTTTTACAAATTCCTMAATCAKTTAATAATFACATTTACAAAAMAATGTA 577 QY 280 caaagctatgtgcgatcgtatgagccctctatggtcccaagaatcggatgacattaa 339 DB 578 AATATCTCKTTMAACAAANTCTTATGTCACCTACAGTCAAAATAATKATCTKYAAMCCCMWT 637 QY 340 agtatctataactaccctaca 360 DB 638 KTKTKTMAACTCCACTTAA 658 </pre>	
RESULT 13	AO368869	391 bp DNA GSS 06-MAR-1999
LOCUS	AO368869/c	
DEFINITION	HS_5021_B2_C07_SF6E RPC111 Human Male BAC Library Homo sapiens genomic clone Plate=597 COL=14 Row=F, DNA sequence.	
ACCESSION	AO368869	
VERSION	AO368869.1	GI:4338348
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu	

Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 597 row: F column: 14

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 391.

Location/Qualifiers

FEATURES

source

1. 391

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate:597 Col=14 Row=F"

/clone_lib="RPC11 Human Male BAC library"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACE3.6; RPC11 Human Male BAC Library"

/note="Vector: pBACE3.6; 106 t 1 others

108 a 93 c

83 g

106 t

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

1 others

Query Match

Best Local Similarity 63.08; Pred. No. 10; Length 391;

Matches 51; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 1 cgaatggtcgtatgtgttcacagagccactaacctataatggtgactagatgct 60

Db 226 CTATGCGCTCGCTGTGCGGGAGGCTGAGGAGCTCCACTAGCATGATCTTGATGCT 167

Qy 61 ctgggcaaatcaatgacc 81

Db 166 CAGGCGCAATTCATGTCGCC 146

RESULT 14

AZ618773

LOCUS

1M0450H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

ACCESSION

AZ618773

VERSION

AZ618773.1

KEYWORDS

SSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. 406

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 406.

Location/Qualifiers

1. 406

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 406.

Location/Qualifiers

1. 406

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 406.

Location/Qualifiers

1. 406

/organism="Mus musculus"

/lab host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: pMD29v; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pMD42 (g147321149pAF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance.

Qy 182 ttctgtagtggggcttcgttcacagcagtcactcgtatgtagtag 241

Db 180 TTTCAGTTCTTGAGATTCTGTTATATGAGAGCTGCTGCTGCTGCTGCTG 239

Qy 242 catatactgctgctgctcactcgtatgtagtagcactcaagctatgtagtag 299

Db 240 CTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 289

Qy 300 taggcctcctatgct 314

Db 300 CTAGTCTGCTATGCT 314

RESULT 15

AZ687120

LOCUS

ENTGA85TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

ACCESSION

AZ687120

VERSION

AZ687120.1

KEYWORDS

SSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. 869

/organism="Entamoeba histolytica"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 816.

Location/Qualifiers

1. 869

/organism="Entamoeba histolytica"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 816.

Location/Qualifiers

1. 869

/organism="Entamoeba histolytica"

/db_xref="taxon:10090"

/clone="UUGC1M0450H04"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

High quality sequence stop: 816.

Location/Qualifiers

1. 869

/organism="Entamoeba histolytica"

```
/strain="HML:IMSS"  
/db_xref="taxon:5759"  
/clone_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."  
BASE COUNT      273 a      132 c      115 g      349 t  
ORIGIN
```

```
Query Match          9.0%; Score 33; DB 225; Length 869;  
Best Local Similarity 47.8%; Pred. No. 13;  
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
```

```
QY 101 tgtgcttgygacttagaactgctgcgtaagtaagtaaggaagtcacttctca 160  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 107 tgttatatttatatagtcatttgatattgagttacttttctaactttcatcttgat 166  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 161 ctacaatatatacactcgaattcgtgattggggcttcgttcataagcagtcgaact 220  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 167 ctatataacacacaaaataatggtttatttggccttttggtttaacggttattactc 226  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 221 actctgactgactcgaatcattactactgctcgctactcgaactcgaatcgaatcgaatc 280  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 227 catctatttttttaagactagcagtgaaatgctcagagatttgatttaattgattatctatat 286  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
QY 281 aaagcttatgcatgcatca 301  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 287 tattcatatataataataatga 307  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Search completed: May 8, 2001, 05:34:30  
Job time: 9039 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:55:34 ; Search time 2791.02 Seconds
(without alignments)
2187.592 Million cell updates/sec

Title: US-09-619-643-4

Sequence: 414
1 cggacgcgtggcgagcg.....gtgttttgcgtgacctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rtd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_roi:*
95: gb_roi2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66.8	16.1	1455	94	AF142630 Mus muscu
2	64.4	15.6	1629	94	AF142629 Rattus no
3	62.2	15.0	1492	4	AF002677 Dictyoste
4	58.8	14.2	1743	88	AF155140 Homo sap1
5	58.6	14.2	2552	94	MUSHERATP
6	54.2	13.1	36149	14	SPBC12C2
7	54.2	13.1	92611	13	ATT4D2
8	53.4	12.9	43460	63	AC014216
9	53.4	12.9	67242	61	AC009539
10	53.4	12.9	234138	4	AE003678 Drosophi1
11	53	12.8	3092	8	AF190623 Xenopus 1


```
Oy 93 cgaggttttaccacaacacgtaacctgtcatcaactacacagatgcgaatcaattt 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1243 CGAGGATTTGATGATGACGACGATGACAAATTTGTTGAACTTTGACCTCCCTGTAACAA 1302
Oy 153 ggtacaagaagatgaacccgattatagtgctctccacagatggcagaagctggcg 212
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1303 G-----GAGAGAGACCGCGACTGTAGACCTTCCTCCGACCGATAGGCGGAGCGGCGC 1356
Oy 213 tttagcggaagaagctgctgtctcaacttgc 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1357 TTTGGGAAAAAAGCGCTTCCTCCCTCAACATGAT 1388

RESULT 5
MUSHELATP 2552 bp mRNA ROD 25-NOV-1994
LOCUS Mouse RNA helicase and RNA-dependent ATPase from the DEAD box
DEFINITION family mRNA, complete cds.
ACCESSION L25125
VERSION L25125.1 GI:407991
KEYWORDS ATPase; DEAD box family; helicase.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2552)
AUTHORS Gee,S.L. and Conboy,J.G.
TITLES Mouse erythroid cells express multiple putative RNA helicase genes
JOURNAL exhibiting high sequence conservation from yeast to mammals
MEDLINE Gene 140 (2), 171-177 (1994)
FEATURES
SOURCE location/Qualifiers
1..2552
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="erythroleukemia"
98..1534
/note="putative: RNA-dependent ATPase"
/product="RNA helicase"
/protein_id="AAA53629.1"
/db_xref="GI:407992"
/translation="MARDSMALAVDEQPAVKSSDIOIKEKAKSDTNGYIKSTTA
EKTEEEKEDRAQSLNKLIRSNLVDFTNVEVLQRPSPPLSVSFEELRLKPOL
LQGYVAMGFNRPSTIOENALPMLAEPPQNLIAQOSCTGTAAVLAFLMRLVPADR
YPQICLSPTVELALQTKGVYEQMGKEFHEKLAVAVGNLREGOKVSEQIVGTG
TVLDMCSKRLKIDPKIKFVLEADAVIATOGHODOSIRIORTVPRNCOMLFSATF
EDSVWFAOKVVPDPIITIKIKRREETLOTIKOYVLCNNRREKFOALCNLTGATTTD
AMTCHRTKASMLAAELSKGHOVALISGEMVQORAVYERERREGKYLVTNTVC
AKGIDVEQSVVINFDELVDKGNPDNETYLHRTGRGRFGKRGGLAVNMVDSKHSMT
LNRIQEHNRKIERLDVDDDEIKIAN"

BASE COUNT 673 a 630 c 653 g 596 t
ORIGIN

Query Match 14.2%; Score 58.6; DB 94; Length 2552;
Best Local Similarity 53.0%; Pred. NO. 1.1e-06;
Matches 152; Conservative 0; Mismatches 129; Indels 6; Gaps 1;
```

```
Db 1372 CATTGGCGCCACTGGCCCTTTGGCAAGAGGGCTGTGGCGGAGCAATGTGGACAGCAA 1431
Oy 254 aacgattatgtgtgtagaagaagatcgaaacatttcacagaca 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1432 GCACAGCATGATATTCCTACACAGATCCAGAGCATTTTAATAAGA 1478

RESULT 6
SPBC12C2/c 36149 bp DNA PLN 03-NOV-1999
LOCUS S.pombe chromosome II cosmid c12c2.
DEFINITION 254140
ACCESSION 254140.2 GI:6249529
VERSION 254140.2
KEYWORDS ATP-dependent RNA helicase; dynamin; fnx1;
glucosamine-fructose-6-phosphate; lactoylglutathione lyase; MDR
protein; oxidoreductase; paired amphipathic helix protein;
putrescine aminopropyltransferase; signal transduction; spermidine
synthase; starvation response; transmembrane protein; vacuolar
sorting protein.
SOURCE Schizosaccharomyces pombe
ORGANISM Schizosaccharomyces pombe
REFERENCE 1 (bases 1 to 36149)
AUTHORS Devlin,K. and Churchner,C.M.
TITLES Unpublished
JOURNAL 2 (bases 1 to 36149)
AUTHORS Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
TITLES Submitted (07-SEP-1995) Schizosaccharomyces pombe chromosome I
JOURNAL sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridg
MEDLINE eCB10 IRQ E-mail: barrell@sanger.ac.uk Rieger M., Biotechnologische
FEATURES und molekularbiologische Forschung, Angelfhofweg 39, D-69259
Wilmshelmfeld, Germany
COMMENT On Nov 4, 1999 this sequence version replaced gi:984208.
Notes:
02-OCT-1995 revised splice pattern for SPBC12C2.11 20-MAY-98
corrected initial chromosome I assignment changed all SPAC
designations to SPBC accordingly. Added author, Rieger M.
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/
yeastpub/svw/home.html)
protein coding regions (CDS) have been predicted with the help of
computer analysis using the GeneFINDER program in Pombase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of Introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC5H10.01c. SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
c12c2 is entirely contained within cosmid c17f3.
However for submission purposes c17f3 has been cut to nominal w.r.t
c12c2.

FEATURES
source location/Qualifiers
1..36149
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/clone="cosmid c12c2"
/map="11"
1..569
gene
```

CDs

/gene="SPBC12C2.13c"
/note="fnx1; SPBC21D10.04c"
1..569
/partial
/gene="SPBC12C2.13c"
/codon_start=3
/product="MDR protein"
/protein_id="CA90826.1"
/db_xref="GI:984220"
/db_xref="SWISS-PROT:Q09752"
/translation="SICSGLYMKTKGRRLVYAGFFPMIMGIVSAVLTSFGCHRPPL
ILISLCIAMGSGYGMNLVSTLAIILISSLAPEQAAVATGSLYFRATGVSIGLSQST
TSLIIMKGLAMKDDPKDDKDLIRRLRESISITIPMLPKDIOQLVIRKVSATARTWTFA
LVAILAFAGWCGLRIKQFYLHSTVDS"
complement(1..9176)
/note="Overlap with SPBC21D10 EM:AL031356 S. pombe
chromosome 2"
1343..2251
/gene="SPBC12C2.12c"
1343..2251
/gene="SPBC12C2.12c"
/note="SPBC12C2.12c, len:301, SIMILARITY: Saccharomyces
cerevisiae, LGUL_YEAST, lactoylgultathione lyase, (326
aa), fasta scores: opt: 951, E():0, (49.8% identity in 305
aa)"
/codon_start=1
/label="SPBC12C2.12c"
/product="lactoylgultathione lyase"
/protein_id="CA90825.1"
/db_xref="GI:984219"
/db_xref="SWISS-PROT:Q09751"
/translation="MASTDMSYTKLNTMIRYKDLKSLKFEYEGMKLIDOWFE
ENEFSSFLAFQPCALNHCVERSKREGILELYNCTEKREBPVYINGTEKRRFG
HICFTYDNIESACATYLESKGVSRKKLSDCKMHIAPADPDWYIELVQSSTKRA
NISNFRNHTMVAVKDPEPIAYEKLGMKVIDKADHPNGKFTNYFLAIPSDLPDRD
EGLEJLHMGTGEEKSGPYVHNGDGDDEKQGHVCIIVNINAACSKFEAEGLPFKKK
LEGRMKDIAFLDPDNYWEYIEQK"
1394..1801
/gene="SPBC12C2.12c"
/note="Match to PF00903 Glyoxalase, Glyoxalase Score
302.20"
1859..2245
/gene="SPBC12C2.12c"
/note="Match to PF00903 Glyoxalase, Glyoxalase Score
301.26"
2006..2035
/gene="SPBC12C2.12c"
/note="PS00934 Glyoxalase I signature 1"
complement(join(3136..5150,5356..5426,5495..5499))
/gene="SPBC12C2.11"
complement(join(3136..5150,5356..5426,5495..5499))
/gene="SPBC12C2.11"
/note="SPBC12C2.11, len:695, SIMILARITY: Candida albicans,
GFL1_CANAL, glucosamine-6-phosphate
aminotransferase [isomerizing] (ec 2.6.1.16), (712 aa),
fasta scores: opt: 2501, E():0, (58.5% identity in 716
aa)"
/codon_start=1
/label="SPBC12C2.11"
/product="glucosamine-6-phosphate
aminotransferase"
/protein_id="CA90824.1"
/db_xref="GI:1006748"
/db_xref="SWISS-PROT:Q09740"
/translation="MCGIFQYINLYVERDRGTYLTKLVGKLEKRGYDSSGCAVDG
DEGEDIMFKREVGNVSKLEASIGSNVNSSTKINICALISHTMAHGIJSPINCHQ
RSDPHSEFIVNGILTNRELRTVLESROMVSESTDECVAKLFIYDTGVGE
TSLAKVFELEGAVALIKTSISHPGEVATRGSPLYVCGEOKLAKDFVEVEPE
PAEGLPQPKPTSLHPVSNPATNGMLRGDKPLLRASRARVSGSGVGPTEVEFA
SDATPIETETKRYMLLEDDIDIAVRDGLVHLHRRGGSGSTRITETLEMETASVAK
GNDRHMIKICQDPSSLMTMRGRVNFVRLTGLGLESYDITIRSRRLIVACGT
SYHSVAVRPLFEELNIPVVELASDFVDRCSVDRDPIFVSSGSEFADSLRAGT
YILENGALIGVNCVGSISIRKTHGCHINAGPEICVASTKAYTSQYVALVIMALYL

misc_feature

/gene="SPBC12C2.11"
/note="Match to PF01380 SIS, SIS domain Score 139.50"
complement(3697..4098)
/gene="SPBC12C2.11"
/note="Match to PF01380 SIS, SIS domain Score 161.02"
complement(join(4627..5156,5356..5425))
/note="Match to PF00310 GATase-2, glutamine
amidotransferases class-II Score 347.63"
complement(5151..5165)
/note="ctaatcaggttag, splice branch and acceptor"
complement(5350..5355)
/note="gtaaat, splice donor sequence"
complement(5427..5441)
/note="ctaatcagagctag, splice branch and acceptor"
complement(5489..5494)
/note="gtaagc, splice donor sequence"
join(6697..6826,6964..11402)
/gene="SPBC12C2.10c"
/note="SPBC12C2.10c, len:1521,
SIMILARITY: Schizosaccharomyces pombe, O13919, hypothetical
124.8 kd protein c23c11.15 in chromosome 1., (1075 aa),
fasta scores: opt: 1282, E():0, (27.8% identity in 1167
aa)"
/codon_start=1
/label="SPBC12C2.10c"
/product="transcriptional regulatory protein"
/protein_id="CA90823.1"
/db_xref="GI:984217"
/db_xref="SWISS-PROT:Q09750"
/translation="MAKMODARLGGCHREDEYSNVAINTGYLPISGTMAYHPGNA
PLFTQAPRHTNPGCPPEPLNSISPYVDATGRLLYRNNTQVSHRALIPNANGAA
VYGPSPQDPPPPQSHPNVYISAPARALIQPTLISSTDSINIPRETVSSASAI
EVPNPNPSAPPPOEYRLQNLVADSLVLOHFOPEIYNEFLDIKERKSOAI
ETPEVITRVSKLFAQYVNLIOGFNTFLPPQYSISADPGSLAGIHITTPQPSLAP
DLGKTTAPPPHGSTPLPAASATYSMNKQSSASHPVLOPAPSTLOEGLSPAPAP
SYPPDASVYKQADLDQALINFNVNNKRRSEKAEANSFELDKSYOHDPPIQLYF
QVSLPAEAPDLLEFKRRLPDVSNVNAPEQOKSVYQESATATPKSPATPSA
LPPIGKFAPTTAKAOPAPKRRGEPAVOTRNKSKRTTRTSVEETTPRAVNPVIAO
NKNPSELEFLEHARQYLANESKYNEFKLELVSQVFPKNAALVECYVFFGSEHILM
NMLKDLVKNPNAMPVYPRPVDLTQCKSGSPRYRLPKTEILLPQSGRDLQWTILN
DAVWSEPTLASDESGFLAHRKNQFEENLLEERYEDRHIKANRFEILQIHADK
MLKMSVEKAMNTLPSNLGKSVSIYHKVYKYGKHAQOIIENLOKNSVTPVYL
ERLKKDRERSLOHNMELMHDIIEKNFERSLDHOGVSKSVYDKSTPKELISLR
NLAQOKVELSEKVPVPSHOFLSYKDPNITIDTARLFGVFLIHGSTHAKEDNEKSN
FLRSFLSLEFDPYDSFTPLYPHFNEESDDISLSLLEKRASSSPHANNNGL
RLKADVLKTYRGARERNSVSKEDYVSESERPPDASEIDEHISEHENDDESSVYS
TGEVWVCKTDDGSLDDGTCLDSRYNLKGNMSLVCFFLPFTLVLSRLSEIKNL
EOMAYKOHQVKNPVAVELGVRHPSERGLFPADVYEAQALQCELMAGREDO
NGEEDALRCYLGHAFLRYVEKLYSIIKLOHSYTNRLAOPVYKEDMDRORRST
PROQINRIOTETAFGPDENLCCIDMNSQTSRATILMGRREDLTKGLSDAKKCY
IGSYIMSSPTGEGILPEHVRILPFLRKCLPSDEGNEDESSVVAANAIIITSFLESLA
LTPIPTVYKTRYENGTEDEVARNSEOVYNGPYKIDYRQSKREMLNSDEGTGILS
KDRVRIKIPCTIESTLNESTLRSGKAREFSENGSIVIGKKGSEIDEEKNGKLLG
LPPVNGKSSVYTGNTNLRKARNNDNSNINISSEKKEESTIEDEEKNGEKS
PVAKHASDVEDHDVAKSTAPDETSVSHRPERSEKKSPPSVYTSVKQTAENDADND
DKTDMDDOTETETLADADNTMEERSKDL"
6827..6832

misc_feature

13.1%; Score 54.2; DB 14; Length 36149;
Best Local Similarity 53.2%; Pred. No. 3.4e-05;
Matches 141; Conservative 0; Mismatches 118; Indels 6; Gaps 1;

QY 33 ataatacagaatcaagaatgctacacggaggtcttatatacaccagatgtcttgc 92
11
Db 19324 ATTATGATTCATCATCCAGATGAGTACTCAAAAGGTTTGGTTACTACGAATGTAATTGCT 19265

/number=7
6885..6996
/gene="T4D2..20"
/number=7
6997..7249
/gene="T4D2..20"
/number=8
7250..7358
/gene="T4D2..20"
/number=8
7359..7529
/gene="T4D2..20"
/number=9
7530..7644
/gene="T4D2..20"
/number=9
7645..7737
/gene="T4D2..20"
/number=10
7738..7815
/gene="T4D2..20"
/number=10
7816..7923
/gene="T4D2..20"
/number=11
7924..8047
/gene="T4D2..20"
/number=11
8048..8164
/gene="T4D2..20"
/number=12
8165..8256
/gene="T4D2..20"
/number=12
8257..8346
/gene="T4D2..20"
/number=13
8347..8437
/gene="T4D2..20"
/number=13
8438..8521
/gene="T4D2..20"
/number=14
complement(9057..9151)
/number=1
9057..10601
/gene="T4D2..30"
complement(join(9057..9151,9403..9658,9810..10043,
10124..10254,10370..10601))
/note="strong similarity to proline-rich protein, Raphanus
sativa, PIR:S16748"
/codon_start=1
/product="putative protein"
/protein_id="CAB64213.1"
/db_xref="GI:6630730"
/translation="MRVSGPRVLLVSCFCKSKGAVPALIMFGSDIYVGNNNLL
STYKSNFLPYGRDFTDQRTGRCNKRLAVDSAEYLGSSYTPATLSREASENTLIT
GANFASAGGYDATSVPGSISLTROLSYRAYQNRVIMIGRNARILIFSCHIL
SAGSSDFLONYINPLNLTNPDDADILRSFSEIQLNYELGARIGVLSIPMG
CLPAATTLFGAGKSCVERLNDAMFMENTYRTLLMRHSGRLVAFNVYQPLD
ITNPTDNGYVEWDGHPTEAVNELLAGQLLGGISLIN"
complement(9152..9402)
/number=1
complement(9403..9658)
/number=2
complement(9659..9809)
/number=2
complement(9810..10043)
/number=3
complement(10044..10123)
/number=3
complement(10124..10254)
/number=4

Query Match 13.1%; Score 54.2; DB 13; Length 92611;
Best Local Similarity 62.6%; Pred. No. 3.9e-05;
Matches 102; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 86 tctgtcgcgaggttttaccacagacacgttaacctgtcatcaacacagacatccaat 145
Db 14031 TCTTGTGGTTATATACAACTCTGCAGTGGAATTTGGTGTCAATATATCTCCAC 14090
QY 146 caaatttgatcaagagatgaacctgattatgaggttactgtcacagaaatggcagac 205
Db 14091 TAAATATGAACACTGCA---GAGCCAGATTATGAGGTACTCTTACACAGCGTTGGAGAGC 14147
QY 206 tggcgcttggccggaaggtgtgtgtcaactgtgt 248
Db 14148 TGCCGGTTGTGTCGCAAGGTGAGTTTGTATCTCTTTT 14190

RESULT 8
AC014216/c 43460 bp DNA HTG 16-NOV-1999
LOCUS
DEFINITION
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014216 GI:6437119
VERSION AC014216.1
KEYWORDS HTG; HTGS_PHASE2.
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 43460)
REFERENCE
AUTHORS
Adams,M. and Venter,J.C.
TITLE
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
JOURNAL
Rockville, MD, USA
COMMENT
This sequence was identified as CDW:10212845 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..43460
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 11655 a 10031 c 9929 g 11845 t
ORIGIN

Query Match 12.9%; Score 53.4; DB 63; Length 43460;
Best Local Similarity 60.8%; Pred. No. 6e-05;
Matches 87; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 11 ggcgagcgcggtggggaagaataatcaatgaatgaatgaatgctcaaccggggtct 70
Db 37646 GCCCAAAAGAGGCTGACGATCATGAMGAGATCTCCAGCCGCGACGAGTCT 37587
QY 71 tatacaacgatgttctgtcgcgaggtttgacacagcacagcttaacctgtcatcaa 130
Db 37586 CATACACACGATGTGGGCTCGGGGATGAGTACAGAGAGTTCGGTGGTCATCAA 37527
QY 131 ctacgacatgccaatcaaatgtg 153
Db 37526 CTACGATTGGCCCAACACCGTG 37504

RESULT 9
AC009539/c 67242 bp DNA HTG 12-OCT-1999
LOCUS
DEFINITION
Drosophila melanogaster chromosome 3 clone BACR33N15 (D828) RPT-98
33.N.15 map 85A-85A strain Y; cn bw sp. *** SEQUENCING IN PROGRESS
***, 72 unordered pieces.

ACCESSION	AC009539	
VERSION	AC009539.5	GI:6041739
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 67242)	
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Solt,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.	
TITLE	Sequencing of Drosophila melanogaster	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 67242)	
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Solt,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (27-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA	
COMMENT	On Oct 15, 1999 this sequence version replaced gi.6041722. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bugdef@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases. * NOTE: This is a 'working draft' sequence. It currently * consists of 72 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.	
	1	857: contig of 857 bp in length
	858	937: gap of unknown length
	938	1879: contig of 942 bp in length
	1880	1959: gap of unknown length
	1960	2458: contig of 499 bp in length
	2459	2538: gap of unknown length
	2539	3454: contig of 916 bp in length
	3455	3534: gap of unknown length
	3535	4141: contig of 607 bp in length
	4142	4221: gap of unknown length
	4222	5238: contig of 1017 bp in length
	5239	5318: gap of unknown length
	5319	6060: contig of 742 bp in length
	6061	6140: gap of unknown length
	6141	6985: contig of 845 bp in length
	6986	7065: gap of unknown length
	7066	7878: contig of 813 bp in length
	7879	7958: gap of unknown length
	7959	8563: contig of 605 bp in length
	8564	8643: gap of unknown length
	8644	9615: contig of 972 bp in length
	9616	9695: gap of unknown length
	9696	10727: contig of 1032 bp in length
	10728	10807: gap of unknown length
	10808	11727: contig of 920 bp in length
	11728	11807: gap of unknown length

```

* 50791 50870: gap of unknown length
* 50871 51550: contig of 680 bp in length
* 51551 51630: gap of unknown length
* 51631 52342: contig of 712 bp in length
* 52343 52422: gap of unknown length
* 52423 53019: contig of 597 bp in length
* 53020 53099: gap of unknown length
* 53100 53785: contig of 686 bp in length
* 53786 53865: gap of unknown length
* 53866 54491: contig of 626 bp in length
* 54492 54571: gap of unknown length
* 54572 55204: contig of 633 bp in length
* 55205 55285: gap of unknown length
* 55285 56073: contig of 789 bp in length
* 56074 56153: gap of unknown length
* 56154 56873: contig of 722 bp in length
* 56876 56955: gap of unknown length
* 56956 57557: contig of 601 bp in length
* 57557 57636: gap of unknown length
* 57637 58220: contig of 584 bp in length
* 58221 58300: gap of unknown length
* 58301 58926: contig of 626 bp in length
* 58927 59006: gap of unknown length
* 59007 59636: contig of 630 bp in length
* 59637 59716: gap of unknown length
* 59717 60399: contig of 683 bp in length
* 60400 60479: gap of unknown length
* 60480 61187: contig of 708 bp in length
* 61188 61267: gap of unknown length
* 61268 61925: contig of 658 bp in length
* 61926 62005: gap of unknown length
* 62006 62707: contig of 702 bp in length
* 62708 62787: gap of unknown length
* 62788 63495: contig of 708 bp in length
* 63496 63575: gap of unknown length
* 63576 64307: contig of 732 bp in length
* 64308 64387: gap of unknown length
* 64388 65030: contig of 643 bp in length
* 65031 65110: gap of unknown length
* 65111 65777: contig of 667 bp in length
* 65778 65857: gap of unknown length
* 65858 66516: contig of 659 bp in length
* 66517 66596: gap of unknown length
* 66597 67242: contig of 646 bp in length.

```

FEATURES:

Source

```

1..67242
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"

```

Query Match 12.94; Score 53.4; DB 61; Length 67242;

Best Local Similarity 60.88; Pred. No. 6.5e-05; Matches 87; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

OY 11 ggcgcagcgtgagggaagaataacacaggaatcaaaagatgcacaccggagttct 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15154 GCCCAAAAGAGCGGTACGATCATGAAAGAGTCCGAGCCGCGCATGCGCAGTCT 15095

OY 71 tatataaccgaatgtcttgcgtgaggtttgacaaacagtcacactgtcatcaaa 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15094 CATCACACCGATGTGCGTCTGGGTATGATGATGATGATGATGATGATGATGAT 15035

OY 131 ctacgacatgcacataaatitg 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15034 CTACGATTTCGCCCAACAACCTG 15012

```

RESULT 10
AE003678 234138 bp DNA INV 05-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386035 section 3
DEFINITION of 105, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AE003678 AE002708
AE003678.2 GI:10726379
HTG.

Fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 234138)

REFERENCE
AUTHORS

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,
Brannon,R.C., Rogers,Y.H., Blazet,R.G., Champe,M., Pfeiffer,B.D.,
Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abail,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cagle,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferlita,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
Gelbart,M.M., Glasser,K., Glodek,A., Gong,F., Gottrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ideyama,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kensinon,S.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mallet,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Mitsuhashi,N.V., Mobarry,C.,
Morris,J., Mosnig,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusker,D.R., Paclet,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,S.R., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Syrakas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 234138)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7298993.
Location/Qualifiers
1..234138
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
18413..>20074
/gene="CG7800"
/product="CT3295"
/db_xref="FLYBASE:FBan0007800"
/db_xref="FLYBASE:FBgn0037552"
/db_xref="FLYBASE:FBgn0037552"
18413..>20074
/gene="CG7800"
/map="84F3-84F4"
/db_xref="FLYBASE:FBan0007800"
/db_xref="FLYBASE:FBgn0037552"
18473..20074
/gene="CG7800"

/note="CG7800 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0007800"
 /db_xref="FLYBASE:FBgn0037552"
 /protein_id="AAF54196.2"
 /db_xref="GI:10726380"
 /translation="MYLIITALLASVSAAGRPDEDYCNDSYCHLLCRSTPSDKAKT
 KLTHERHMOSEKVKYLAIMPNTLTLELECSDPDTMDLNDQPLVTSIQLRKGLGL
 HDHRSKPNMKIMLGNNTIRTSNCEFKIADLMLSLPGNSIQGLPMDVPCNPE
 LHLDSNRNIEETLHENIEFTGVPKLEMLLNGNPLWTAPISLKSILNRLLDMSNG
 PLPDLISGAHPLILDNSGVOIRDLSEITISITYAAASDNHTESEIKLPDKSSYLEIDHSN
 LITADIDIKLGTGMRRLODRDLSENITISITYAAASDNHTESEIFILNPMYANRNRT
 RLHSPISIPWELETHLDASYNRTIAPAAVGVIDEAFNQSLEBSNYINNELELPKPH
 PSLKVALVYDNCKEOPQKRYKNITKEFNEGAVLLEKTYOSQSNNTTPCKQIPADPF
 PPSIADNPQNTIDTGVNPLSDNDTYOMNVMNVLMLSVLSFNSFLVTKIIRLGRGN
 QPTOSSVPAAITKEMSNESDDEILL"
 20581..22277
 /gene="CG18249"
 /product="Ct41336"
 /db_xref="FLYBASE:FBan0018249"
 /db_xref="FLYBASE:FBgn0037553"
 <20581..>22277
 /gene="CG18249"
 /map="84P3-84P4"
 /db_xref="FLYBASE:FBan0018249"
 /db_xref="FLYBASE:FBgn0037553"
 20610..22277
 /gene="CG18249"
 /note="CG18249 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0018249"
 /db_xref="FLYBASE:FBgn0037553"
 /protein_id="AAF54197.1"
 /db_xref="GI:7298995"
 /translation="MHPKYTLILGCGITVSADALAYDASGRFLUSACDPDRCSLSD
 RFAVASPAALDRLRLHRLNCSROSTIMVLVQLTPGLKTLVTRNCATYHLSKESLRG
 ENUVLSQMGQSLGVLRLDQIFNAVPAEPILOLSONFIHYVAAPGSLSLKRLSDLV
 NAIATILSSTDPLMELVLDLSRNEITLTPQNFPAKKIKQTLILNGNPRIILMPYV
 LGSILNRLDLGHALELVMTLNTQONLVLEGSSLSLVINGCFITKQAGNNEANL
 HLOVNRKSSTVEMDHLGNLNGNDPAALRGCMNLORLDSKRNITMALPHGSSLDAS
 GHOELLIPSKIKFLNANNOLVNPRLPESPISTSLSYLDLSHNMILDLVALISLVS
 LKGLYEGNRLLATITYOKLHEHPDLSEGLAHDPWMSGLTRKKEFLFTDGHNLAR
 PGNRVNLNCSRVLDIMPNEQAEQKMDPEGVGIPHVWLRDILAFVTLVVMILL
 LMNVLHLEEBCECLRRFRMRSRITLGQNTSTTRNSGRRLDEQDSEV"
 join(34005..34058,34663..35634,35699..>36955)
 /gene="CG7602"
 /product="Cm23191"
 /db_xref="FLYBASE:FBan0007602"
 /db_xref="FLYBASE:FBgn0037554"
 <34005..>36955
 /gene="CG7602"
 /map="84P4-84P5"
 /db_xref="FLYBASE:FBan0007602"
 /db_xref="FLYBASE:FBgn0037554"
 join(34678..35634,35699..>36955)
 /gene="CG7602"
 /note="CG7602 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0007602"
 /db_xref="FLYBASE:FBgn0037554"
 /protein_id="AAF54198.1"
 /db_xref="GI:7298996"
 /translation="MDPASYVLGKSEBAHORTIILHDMQYFYNAOYEIRDPTRLKALGI
 OKNTLVGCNCFYARAKGVTKMLIAEOROTPDILVNGEDLAPYRMSQRIDILIN
 YTPLYVEKGFDEBNHDVATVLELRQAHYAEALLRPVGHNTPADGTPLSNDCGCCAR
 LATIGRIAOIEEELKRLGITCCAGIAYNNKLIVKSSHEPQOQVAVSTYAEQM
 RELGLKPRYTGQKTOCLILEAGMSVYEQLOQCDMDVMRKGFETATRIARLDSEY
 DTSILVRPGKPTIGMEDACKPIIVRTIPEERFPMMLKRLVYQVAAEDGVRPIAKVYL
 RKFDQSKSSHETKOANITLSPLFTSKSPEGTGVSQVLADAGQDKLITVNRLEPR
 IYDAKPNENITIGLAEPSKROBKYGSSINFLIKKADLEVGYSITSLTNSLTSPIA
 ESPYDECAFRSPPTTERPSDOQTRRRRTATSPVMLDNGSESAANSDSESETE
 VESPKRSAGLIVLSKRSRLAADGVGSAAVASPSKLYACDLNLRNDEKDFPMKT
 TPSTSTSAIPAEFRVQPPNTLQRIODSLFVFTTPTRASRLSSNASTASPSLPSMD

[illegible]

ACCESSION AF190623
 KEYWORD AF190623.1 GI:9801986
 SOURCE
 ORGANISM African clawed frog.
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 3092)
 AUTHORS MacArthur,H., Houston,D.W., Bubunenko,M., Mosquera,L. and King,M.L.
 TITLE DEADSouth is a germ plasm specific DEAD-box RNA helicase in xenopus
 related to elf4A
 JOURNAL Mech. Dev. 95 (1-2), 291-295 (2000)
 MEDLINE 20368186
 REFERENCE 2 (bases 1 to 3092)
 AUTHORS MacArthur,H.C., Houston,D.W., Bubunenko,M. and King,M.L.
 TITLE Direct Submission
 JOURNAL Submitted (27-SEP-1999) Cell Biology, U. Miami School of Medicine,
 1600 N.W. 10th Ave., Miami, FL 33101, USA
 FEATURES
 source
 1..3092
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /cell_type="oocyte"
 1..3092
 /gene="DEADSouth"
 44..1495
 /gene="DEADSouth"
 /codon_start=1
 /product="DEADSouth RNA helicase"
 /protein_id="AAP99574.1"
 /db_xref="GI:9801987"
 /translation="MAAKFLPRFRSSQAELEDFONNNVVAEGLDPEHETLKGSKS
 RYGDGEDVVRGHIIEDLANSLINKLRLTVDSPHAVVEYLORPTSPLEFKRFEEL
 HKLEELRGVYAMGFNRPSKIOENALPMLADPONTIAOSGSGTGAFAVLMLSR
 VDAAKRTPOCICSPREIALINQGVKVEEMKPCAGIEVYIALGNRPGKSRLEAOI
 VIGRGVYDMGCFKRLITVENISVYFLDEADVNIWQGSBDSVRYRKSMPKSCOMI
 LEFSTPEDSVWAEFAERIVDPDNIILKLEELTLNIDQFYDQCENKQKYSALCNLYG
 VITTAQAIVCQOTRKIASWLSOKLSDGHOVALSGELPYVDRADMIORFREGREKVI
 VTTWVCARGIDVEQVSIYVNEDELIVNDGSDVFETVYLRHIGRTGRFGKKGIASLIEN
 FEVYMLKEIEDHFMTKTKLNSMDMDMKRWK"
 BASE COUNT 933 a 535 c 628 g 996 t
 ORIGIN
 Query Match 12.8%; Score No. 5; DB 8; Length 3092;
 Best Local Similarity 56.9%; Pred. No. 5.3e-05;
 Matches 120; Conservative 0; Mismatches 85; Indels 6; Gaps 1;
 QY 31 agataatacaggaattcaagaatggtctacacggggtctctatatacaacgcatgtctg 90
 Db 1179 ACATGATCAAAAGANTTCGCGTGAAGGAGGAGGAAGATCTTGTACTACTAACGTTTGTG 1238
 QY 91 ctcgaggttttgaccacgaacacgcttaacctgtcatcaactacagacatgccaatcaat 150
 Db 1239 CACGAGGATTGATGATTGTAACAGCGCTCCATTGTGTGTAATTTTGACCTGCCAGTAATMG 1298
 QY 151 tttagacaagaagatgaacctgatatatggtggtgactgtcacagaattggcagagctggc 210
 Db 1299 TTGATGGCTCGATGAC-----TTTGAAACCTACTTGTGATGATGTTGGCCGCACTGGGC 1352
 QY 211 gctttgcgcggaagaagtgctgtgttcaact 241
 Db 1353 GATTTGAAAGAGAGGATTTGACAGTACGACCT 1383
 RESULT 12
 RICEIFA4
 LOCUS RICEIFA4 1549 bp mRNA PLN 03-FEB-1999
 DEFINITION Rice mRNA for eukaryotic Initiation factor 4A, complete cds.
 ACCESSION D12627
 VERSION D12627.1 GI:303843

KEYWORDS	eukaryotic initiation factor 4A.
SOURCE	Oryza sativa (strain Yamahousi) callus. cDNA to mRNA, clone S105.
ORGANISM	Oryza sativa
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
REFERENCE	1 (bases 1 to 1549)
AUTHORS	Nishi,R.
TITLE	Direct Submission
JOURNAL	Submitted (13-JUL-1992) to the DDBJ/EMBL/Genbank databases. Ritsuko Nishi, Hokkaido University, Department of Botany, Faculty of Science; Kita-Ku Kita 17, Nishi 8chome, Sapporo, Hokkaido 060, Japan (Tel.:011-716-2111(ex.2740), Fax:011-746-1512)
REFERENCE	2 (bases 1 to 1549)
AUTHORS	Nishi,R., Kidou,S., Uchihiya,H. and Kato,A.
TITLE	Isolation and characterization of a rice cDNA which encodes the eukaryotic initiation factor 4A
JOURNAL	Biochim. Biophys. Acta 1174 (3), 293-294 (1993)
MEDLINE	93385159
COMMENT	Submitted (13-JUL-1992) to DDBJ by: Ritsuko Nishi Dept. of Botany Faculty of Science Hokkaido University Sapporo, Hokkaido 060 Japan Phone: 011-716-2111 x2740 Fax: 011-746-1512. Location/Qualifiers 1..1549 /organism="Oryza sativa" /strain="Yamahousi" /db_xref="taxon:4530" /dev_stage="Callus" /tissue_type="Callus" 12..1253 /codon_start=1 /product="eukaryotic initiation factor 4A" /protein_id="BA02152.1" /db_xref="GI:303844" /translocation="MAGMAPGSGQFDAAKHYSKMOELLNOGETEFTSYDEVESHFSPDMGLDNLRLGIVAYVGEKPESALIDRGIVPECKGLDIVIOASGTGKTAPFGCIILOC LDRAVVECOALVALPTRELAQIEKKVRALGDYIGVAVHACVGSTSVREDRIITASCYHVAVGTGRVLEMLRKOSLRPDYIKMFVLVDADMLSRGFKDOIYDFOLPSIQQGVAFSTMPPEALEITIRKPMNKRVRLILVRDELTELGIIQOPVNVKEBMDLTLDLYETLATOQVIEVNTQKVVMLDEDKMRGRDHVTASATGGMDONTRDIIMREFSGSSRYVLTTLDLARGLDVQOQVSLVINVDLPQNPENYLHRIGRSRGFKKVAINPVRTDDRMULEDIOEFYNVVEELPANVADLL"
CDS	
BASE COUNT	359 a 310 c 418 g 462 t
ORIGIN	
Query Match	12.6%; Score 52.2; DB 14; Length 1549;
Best Local Similarity	62.8%; Pred. NO. 8.2e-05;
Matches	81; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
Dn	25 gggaagaataatacaagaaatccaagaagtgcctacaccgagggtccttatcacacgatg 84
Dn	964 GGGACATCATCATGAGGAGATTCCGACTCTGGTTCTTCCCGTGTCCTCAATTACACGGACC 1023
Dn	85 ttcttcgtcgaggttttgaccacaagacacacgtaacctgtgcatcaactacagatggcaa 144
Dn	1024 TTCTTGCCCCGGNGATTCATGCACGACAGGTGTCCCTGTCTCATCAACTATGATCTGCCGA 1083
Dn	1084 CCCAACCTG 1092
LOCUS	AB046416 6250 bp DNA PLN 23-JAN-2001
DEFINITION	Oryza sativa eirp4A-1 gene for eukaryotic initiation factor 4A, complete cds.

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Features
AB046416	AB046416.1	GI:12381839		Oryza sativa DNA.					
				Oryza sativa					
				Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.					
				1 (sites)					
				Kato, A., Fujita, S. and Kameda, Y.					
				Isolation and characterization of two genes encoding the eukaryotic initiation factor 4A in rice					
				Unpublished					
				2 (bases 1 to 6250)					
				Kato, A.					
				Direct Submission					
				Submitted (21-JUL-2000) Atsushi Kato, Hokkaido University, Graduate School of Science, Division of Biological Sciences, Kitaku Kitao					
				Mishis, Sapporo 060-0810, Japan (E-mail: atsushiesl.hokudai.ac.jp, tel:81-11-706-2740, Fax:81-11-706-4851)					
				location/Qualifiers					
				1..6250					
				/organism="Oryza sativa"					
				/db_xref="taxon:4530"					
				2463..2652					
				4028..4029					
				join(4030..4100,4210..4316,4401..4831,5242..5874)					
				/gene="eIF4A-1"					
				join(4030..4100,4210..4316,4401..4831,5242..5874)					
				/gene="eIF4A-1"					
				/note="mRNA sequence is reported in Acc# D12627"					
				/codon_start=1					
				/product="eukaryotic initiation factor 4A"					
				/protein_id="BAB1260.1"					
				/db_xref="GI:12381900"					
				/translation="MAGMAPEGSQFDKHKHDSKMOELLNGEPEFTSYDEVHESPD					
				DMGLOEMLRGIVAYGFEKPSAIDORGIPECKGLDVIQAOAGSTGTAAPCGIIIO					
				LDVAVVSCALVLAIPREINAOIEKVRALGDIYGVVHACVSTSYREORIATASV					
				HVYVGTGRVFDMLRROSLRPDIITKMFVLDEADMSRGFRDQIYDFQLPSKIQVG					
				VFSVTPPEALEITTRKPMNKPVRLIVRDELTEGLIQFYVNIKEKEMKIDTGLDIE					
				TLAITQSVIEVNTKRKYDMLDKMRGRDHTVSAHGDMDONTNRIIMEFSGSSRYLI					
				TTDLARGIDVQGSVLIYNDLPTQPEPNYLHRIIGSGRFRKGVAINFVTRDERMLF					
				DIOEFNYVIEELPANAVDLL"					
				5875..6171					
				BASE COUNT					
				1661 a 1273 c 1267 g 2048 t 1 others					
				ORIGIN					
				3'UTR					
				Query Match					
				Best Local Similarity 12.6%; Score 52.2; DB 12; Length 6250;					
				Matches 81; Conservative 0; Mismatches 48; Indels 0; Gaps 0;					
				OY 25 gggaagaataatacaggaaltcaagaatgctacacccggggtctctatcaaccgagt 84					
				Db 5585 gggaacatcatcatgagcagagttccgagatcgcttcccccgtgctcavtaccacgagac 5644					
				OY 85 ttcttgcccgaggttttaccagaacacgcgttaacctgtatataacacgaatgccaa 144					
				Db 5645 ttctttgcccggtggtttatgtatccagcaggggtcccttgatcatcaatgatgctgccga 5704					
				OY 145 tcaaatgt 153					
				Db 5705 cccaacctg 5					

ORGANISM	REFERENCE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Neurospora crassa	Eukaryota: Fungi: Ascomycota: Pezizomycotina: Sordariomycetes: Sordariales: Sordariaceae: Neurospora.					
1 (bases 1 to 46300)	Schulte,U., Algn,V., Hohelsel,J., Brandt,P., Fartmann,B., Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.					
2 (bases 1 to 46300)	German Neurospora genome/project.					
Direct Submission	Submitted (08-FEB-2001) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: mannhauptmips.biochem.mpg.de Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, E-mail: ulrich.schulteuni-duesseldorf.de					
this contig is an assembly of cosmids 8D4 from 1 to: 5127 and 606rev from 5128 to: 46300, cosmid library pLORIST6Xh, strain 74-OR-23-1A cosmid clones are available at the Fungal Genetic Stock Center, http://www.fgsc.net	Sequencing was performed by MWG Biotech AG, Ebersberg, Germany, http://www.mwgda.com					
Information on performance of analysis and a more detailed annotation of this entry and other sequences of linkage groups II and V can be viewed at:	http://www.mips.biochem.mpg.de/proj/neurospora .					
Location/Qualifiers						
1. .46300	/organism="Neurospora crassa"					
/db_xref="taxon:5141"						
/chromosome="2"						
1. .31155	/note="Overlap to Cosmid contig 3H10, EMBL:AL513442, please refer to this entry for analysis and annotation of the first protein"					
3800. .3969	/gene="8D4.20"					
/number=1						
3800. .4613	/gene="8D4.20"					
join(3800. .3969,4111. .4268,4354. .4613)						
/gene="8D4.20"						
/codon_start=1	/product="hypothetical protein"					
/protein_id="CAB88544.1"						
/db_xref="GI:7635778"						
/db_xref="SPFEMBL:09P738"						
/translation="MCRRKMQIEPRSSSTTSVRYRVGCLRHSHVCREGLSSETN OGCTGLDMNRSENRNGRGOGAETGSEGEVKEVEERMGATWTAAADNGESI KCGVRHETFLVACVCGNMYTIRKRAQACVYISGPIPLAIVETSGGLGSGSRSEIV QIQCGPRASCAAAVAVGTCKGGRNSAARSMQRRQT"						
3970. .4110	/gene="8D4.20"					
/number=1						
4111. .4268	/gene="8D4.20"					
/number=2						
4269. .4353	/gene="8D4.20"					
/number=2						
4354. .4613	/gene="8D4.20"					
/number=3						
6155. .7363	/gene="8D4.30"					
complement(6155. .7363)						
/gene="8D4.30"	complement(6155. .7363)					
/gene="8D4.30"						
/note="similarity to hypothetical protein YBR228W, Saccharomyces cerevisiae, PIR:S46104"						
/codon_start=1	/product="conserved hypothetical protein"					
/protein_id="CAB88545.1"						

SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 89795)
AUTHORS Schulte, U., Aign, V., Hohelsel, J., Brandt, P., Fartmann, B.,
Holland, R., Nykatura, G., Nemes, H. W. and Mannhaupt, G.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 89795)
AUTHORS German Neurospora genome project.
TITLE German Neurospora genome, project.
JOURNAL Submitted (07-FEB-2001) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
mannhaupt@mps.blochem.mpg.de Project Coordinator: Ulrich Schulte,
Institute of Biochemistry, Heinrich-Heine-University, D-40225
Duesseldorf, E-mail: ulrich.schulte@uni-duesseldorf.de
This contig is an assembly of cosmid 3H10 from 1 to: 20386, and BAC
8G8 from 20387 to: 89795; cosmid clone (library pLORIST6X), strain
74-OR-23-1A) and BAC clone (strain OR4A) are available at the
Fungal Genetic Stock Center, <http://www.fgsc.net>
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
<http://www.mwgna.com>
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of linkage groups II
and V can be viewed at:
<http://www.mips.blochem.mpg.de/proj/neurospora>.
Location/Qualifiers
FEATURES
source
1. 89795
/organism="Neurospora crassa"
/db_xref="taxon:5141"
/chromosome="2"
misc_feature
1. 9116
/note="overlap to Cosmid X23F10"
CDS
1961..3829
/gene="3H10.10"
/note="similarity to samb, aspergillus nidulans,
TREMBL:ANSMB_1"
/codon_start=1
/product="related to samb"
/protein_id="CAC28640.1"
/db_xref="gi:12718251"
/translation="MREVFESIPNVNKAISVGTITLADRALDCTSLPLNSLHIA
VLTSSARITRLITVDGIERVCLIKGRSDMDMKRNILAFCCVNIIGVETNV
RTTVEADKVPYATILINDYIKVIERVEKAEAKHSDPHHRHISGNRVKKSASF
SSRPINIDVDQAFRRQAQPIDIVSITLAGSTEAAPDETPTTASRFSQILGER
LRFSAHRSHHHHGHVQDOSQAVSPSTRHESQPPAPGSSSEVNDSEAFARFETDR
LRNMGPFANTNLISQPSPTPLPTSVRPSITLAPGSRSPRSIRHOSHAGETDDM
NTDVSADENDDAEMSGTADQNEGYIOQVSEDDGIMTGAALDLYPTASQAEAG
TENTHOGAVDTSMENNAFTPTQVNLGSLSPNPTTANTAQSMPNSRVPHYLDRHT
PNPOMLAMPREDVLMSLQILAVYSKCNLKKYQKCHLVPKLIGKIKLIDGTD
TELEAELEAEDEEEYLLPNDONIFPLVEKFTVRYHSTDMQYAGVVMRLCKRD
TRGGIROCAVYQCGKWEYTRQFAKRCRCRRTKCYKSCQKSAWAFHRHWCAATP"
1961..3829
/gene="3H10.10"
/number=1
exon
1961..3829
/gene="3H10.10"
complement(6485..7221)
/gene="3H10.20"
/number=1
gene
6485..7558
/gene="3H10.20"
complement(join(6485..7221,7293..7338,7400..7558))
/gene="3H10.20"
complement(join(6485..7221,7293..7338,7400..7558))
/gene="3H10.20"
/note="similarity to product: CTNS, homo sapiens,
TREMBL:AF168787_2"
/codon_start=1
/product="related to CTNS protein"
/protein_id="CAC28641.1"
/db_xref="gi:12718252"
/translation="MHVPPLPTSSHPCSTKRRRSRTSIVLAISCPAVLANDGPHFLD

OLIANLMDTSGTTIDPELINSIGFAVTTISNIAFAYSPLIRAQYAVRYKGLPTVQ
LNDITFAEHLICFVTSQYLEPRAMGAPSLGSRPSRLIGTFEGCAGALLIFR
VWASPVPRNTSCAAVTEGGWMLDAIVASVYKLVVTLIKYTPQIVYVNRNSTGMS
ILQILLDEGGGLISTAQOAIDSYOQDMSGIRGNPKALGSVMSIYDLVFLMQHYVL
YRHAGAKADAGNAGSDDDPLRLREYRGGGDEERKIQ"
complement(7222..7292)
/number=1
exon
complement(7293..7338)
/gene="3H10.20"
/number=2
complement(7339..7399)
/number=2
complement(7400..7558)
/gene="3H10.20"
/number=3
8576..10537
/gene="3H10.30"
complement(8576..9478)
/gene="3H10.30"
/number=1
complement(join(8576..9478,9530..10039,10304..10537))
/gene="3H10.30"
complement(join(8576..9478,9530..10039,10304..10537))
/gene="3H10.30"
/note="strong similarity to SRP1 protein,
karyopherin-alpha or importin, YNL189w, Saccharomyces
cerevisiae, PIR:S30884
contains EST db:AI392107, AI330311"
/codon_start=1
/product="probable importin alpha SRP1"
/protein_id="CAC28642.1"
/db_xref="gi:12718253"
/translation="MADRYTPEHRRTOPKASAFAPDELRRRREBOYEIRKAREEN
LAKRRIGAGDSRPGASISGAAPRSDSDDEPPTESQLEDPKMGVSESEIDKOAT
TKFRKLTKERNRPTEBYITKGYVRFEPFLSRPTLVQFEAMALINIASGSRQTO
VIEGAVPATVELIGSPEDVREGDAVVALGNIADSPQCDYVLSGALPLRLTLG
DSRKLMLRNATWLTLSNFCRGKTPQPDWNTIAPALPVLAKLVSLDDEVLIDACMAIS
YLSDSNDRIQAVIEAGIPIRLVELIMHASTSVQPPALRSVGNIVTGDOVOTGVINC
GALPCILLSLSNKGIRKEACWTISNTAGNSAQIOSIVANITIPPLIHLSHADLK
TRKEACMAISNATSGGLQKPOIRYLVAOGCKIKPCDLIAPDKNTIOVALDGENTL
KYGELEDKNAADGSPSINRYALFTECCGMERKIHDCQTNABEIMKRYNITEKFSD
EDEADEMAGQDQGFSGASGAGQGGFNFGANGTESMDM"
complement(9479..9529)
/number=1
complement(9530..10039)
/gene="3H10.30"
/number=2
complement(10040..10303)
/number=2
complement(10304..10537)
/gene="3H10.30"
/number=3
11861..14140
/gene="3H10.40"
11861..14140
/gene="3H10.40"
/number=1
11861..14140
/gene="3H10.40"
/note="similarity to hypothetical protein SPAC26H5.06,
Schizosaccharomyces pombe, PIR:T38425"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAC28643.1"
/db_xref="gi:12718254"
/translation="MMRLDRKPAKAPPLPFSITLPLRAILLDDGDAQPGSMVNV
IGVLKDCRAPATGSDWKCTLTISDSIEDESSAGEVELYIRPAPRGVEVAGDVLV
LSAKVQRKSNPKSLITSKITTCVYKAAITPVYPAQSVALLPAPKGGESHKILKEH
HOVSYLYNVITDKYDVPDEAFYQOQRYVYSLNVKPKSLKAVYTGNTYDLIGOVAKDP
YTDEKRTITLVSYTENDLFFHTTBSGVROLASARPADYTPEDNNPEAQOQHPWG
PYGKRTQIOTSCYDAHADITREAGVSAGWLSLRNVQVFGFNGALHGFLEERNASA
RKVAVELIDIVAPHSVYKAKVDSRLKAVARRMDYSEKKTQILAVYKAQOEACAKK
FARGSSDEVILLDQPPSKKFRAKERRLRAELEKAVEQEKVALGDLNLTVCNAH

```

PKRYSTASILEPYMEYVNGQPVKILPPTNTKYNACVRVVDHPONLEDFACGRK
WSEEDILSDNDHDDDDSDSDSLNSTMDSDNDEDDKGRTEWRFALLIEDASTSL
PATYAAAMDKDKSTKCRDRKTRARLWYVDNTEAQCITGLDANLNRDPPTLTATRE
RMFTLWGNLEELKADKARAAYKREDEGKEDEDEDINGNRTIKRKRKATEVOL
ENLKPLEDSDGE000000HOONOGEEELSNKPFSCCLKOYGVSEDDGLPKTD
GAAGWVRFVGMFGTKICY"
13944..13970
/note="27 bp agcagcagcagcagc tandem repeat"
exon      complement(14521..17508)
          /gene="3H10.50"
          /number=1
          complement(join(14521..17508,17577..17785,17865..18348))
          /gene="3H10.50"
          /gene="3H10.50"
          /gene="3H10.50"
          complement(join(14521..17508,17577..17785,17865..18348))
          /gene="3H10.50"
          /note="similarity to nuclear protein SA-1, Homo sapiens,
          TREMBL:HSSA1.1
          contains ATP/GTP-binding site motif A (P-loop) [GGGDKGS]"
          /codon_start=1
          /product="related to nuclear protein SA-1"
          /protein_id="CAC28644.1"
          /db_xref="GI:12718255"
          /translation="MTTITRQPLMETTNDNTATSSPTSTARRSGRYTKAPKFTPEP
          TSLSKRRKASNDGDEGENSPREGPEDEDDDPASVEYEOBROSRRKRPSSQSAKAK
          KPALKFRKINGDAPSRPOTPAETLASAPALNLSHASLPSRKR SARVAIAQPEGL
          YTDIFRSDDRNEVATRMVHQYQADHAELDLVNCILLSAGCDQRYTEDDIRPDNC
          QNRLADLQNYAEVIGIDYPLISRARSTRFRDLVGFESLVDVLEHEDALYKDVQI
          LENSRRVVASMSSTLRPFRTATTALAIOTALVKAAGIDGRISNVMOOLEKTKG
          GSKNAGLVESIOONLTDABHRYOVCKDOIODIEPTVYVHRQIDAKIRTECEALGN
          MIMLPTVMEPEYLRXGLWMLSDVMPTRHEVYKQGRIFKNNAEKLGHFIDFRFR
          LVENATKDSVNVAVASINQILKDSNLEPDEIDSGRLIFPELIRKAVLDFFA
          GCVNDSEIENKVEEIGEDDAIDELFQDDEDESSPSRDWYSIKCLAEILLAYDQIEE
          ENPEPRNLDIAYDAIPVAPETRI SLAGVLEKEIKIDQKNMELLSGYLLYDHSSA
          SSKSASGGSVETALRAVAPEGREAILLEVLASAVSSLAAPAELEKSKHLRSEG
          IHLGPEEAARKLASIIPKLISKYGADASTAVVLRLEHSLDLNVYOOLGODSTTYDR
          LITGICOFDRHIDRGVLAETTTALLGRKYPALKEITDEKVELMKTINKLRQDK
          SSELGIRNLEEPAILIEGNILLTKSKLYTLKDCIEVLEADQVEGSESPALIEILRI
          VMRKLDQVDENLDPEDAEVAFVKTCLFELMKARVLRDAVSEORSIPALQIERLN
          TLRSFVRNLWTLSSRGTDNDLRLFTGALCDLHVAFAVRYIERNAAGSADYS
          LINPLDPIDGLRELIEFTTAERAVARCIINRLNNGDEGP IEFHSSDEDEDED
          EDDSDLPVOKRKRELKAEVLCETGKMLAIYAKVLDGGSGGAAGKIKRMVVRN
          ONKLGPNKEETLAVLDEGRARAVRBAAKAGKKRPALAAPGSRAGSRPAAAGNS
          GASVAGKKQALSAITTRAEDNSELSDIEIEEEREAVPDEERTDLVNDPIEDHDED
          EGGGDKSKKDDGDEDMDEE"
          complement(17509..17576)
          /number=1
          complement(17577..17785)
intron
exon      complement(17577..17785)

Query Match      12.3%  Score 51:  DB 14:  Length 89795;
Best Local Similarity 61.8%  Pred. No. 0.00035;
Matches 81:  Conservative 0;  Mismatches 50;  Indels 0;  Gaps 0;

Oy 23 999999aaagaataacaggaatcaagaatgctacaccggggtcttatatacaacga 82
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 69738 GAGAGACAGCATTTATGCGAGATTTCAGACAGGCAATACGACAGTCTGATTCGACCGA 69797

Oy 83 tgtcttgcgcgaggttttgaccaagcacacgyltaacctgtcatcaactacgacatgccc 142
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 69798 TGTGTGGGCGCGCGGTATCGATGTCAGCAGGTGAGCTTGTTATCACTATGATCTTCC 69857

Oy 143 aatcaaatgtg 153
      |||  |||  |||
Db 69858 CAGCAACCGTG 69868

```

Search completed: May 8, 2001, 04:57:34
 Job time: 6948 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:13:56 ; Search time 257.37 Seconds
(without alignments)
939.058 Million cell updates/sec

Title: US-09-619-643-4

Perfect score: 414
Sequence: 1 cgcagcgtggcgagcg.....gtgtttgcgcacctctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :
1: N_Geneseq_0401.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	124.6	30.1	1498	21	C39900	Arabidopsis thalia
2	50.4	12.2	765	21	Z88783	T. thermophila Rc2
3	48.8	11.8	486	21	C46718	Zea mays DNA fragm
4	48.8	11.8	496	21	C43533	Zea mays DNA fragm
5	47.4	11.4	1797	21	Z57847	Protein regulating
6	47	11.4	1205	21	F13835	Aspergillus oryzae
7	46.4	11.2	1621	21	C34215	Arabidopsis thalia
8	46.4	11.2	1621	21	C47085	Arabidopsis thalia
9	45.4	11.0	1763	21	C47750	Zea mays DNA fragm
10	45	10.9	1510	21	C38963	Arabidopsis thalia
11	44.6	10.8	737	21	C40626	Arabidopsis thalia

12	44.6	10.8	1082	21	C38774	Arabidopsis thalia
13	44.6	10.8	1636	21	C34340	Arabidopsis thalia
14	44	10.6	1591	21	C44028	Zea mays DNA fragm
15	43.6	10.5	1987	21	C77791	Human cancer assoc
16	43.2	10.4	1575	21	C50339	Arabidopsis thalia
17	43.2	10.4	1577	21	C37681	Arabidopsis thalia
18	40.8	9.9	1052	21	C78203	Human cancer assoc
19	40.8	9.9	1238	15	O57454	Start protein 4A-I
20	40.8	9.9	1383	21	A08808	Androgen-inducible
21	40.6	9.8	1618	16	T05503	Leishmania sp. ant
22	40.6	9.8	1618	18	T80384	DNA encoding lberf
23	40.6	9.8	1618	18	T62139	Leishmania brazili
24	40.6	9.8	1618	19	V47559	Leishmania antigen
25	40.6	9.8	1618	20	X02202	L. braziliensis RI
26	40.6	9.8	1867	18	T80385	DNA encoding lmeif
27	40.4	9.6	2863	19	V38314	Nucleotide sequenc
28	39.8	9.6	235	16	O83249	Streptococcus pneu
29	39.8	9.6	3230	19	V52336	Streptococcus pneu
30	39.6	9.6	897	21	Z53613	Neisseria gonorrhoe
31	39.6	9.6	1275	21	Z53614	Neisseria meningit
32	39.6	9.6	1275	21	Z53615	Neisseria meningit
33	39.6	9.6	3987	21	A81772	N. meningitidis pa
34	39.6	9.6	349980	21	F21610	Neisseria meningit
35	39.2	9.5	1374	21	A70165	Plasmodium falcipa
36	39	9.4	736	21	F11962	Aspergillus oryzae
37	38.4	9.3	33303	21	A81514	N. meningitidis pa
38	38.4	9.3	837096	21	A81489	N. meningitidis pa
39	38.2	9.2	1135	21	C44447	Arabidopsis thalia
40	38.2	9.2	1616	21	C45267	Arabidopsis thalia
41	38.2	9.2	1618	21	C44169	Arabidopsis thalia
42	37.6	9.1	299	20	X84188	DNA encoding human
43	37.6	9.1	299	21	C79417	5' cDNA sequence o
44	37.6	9.1	1830121	17	T42063	Haemophilus influe
45	37.4	9.0	3825	20	X91484	Human C1F30 prote

ALIGNMENTS

RESULT 1	
C39900	standard; DNA; 1498 BP.
ID	C39900;
AC	C39900;
DT	17-OCT-2000 (first entry)
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 26313.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 9905-0121825.
PR	05-MAR-1999; 9905-0123180.
PR	09-MAR-1999; 9905-0123548.
PR	23-MAR-1999; 9905-0125788.
PR	25-MAR-1999; 9905-0126284.
PR	29-MAR-1999; 9905-0126785.
PR	01-APR-1999; 9905-0127462.
PR	06-APR-1999; 9905-0128234.
PR	08-APR-1999; 9905-0128714.
PR	16-APR-1999; 9905-0129845.
PR	19-APR-1999; 9905-0130077.
PR	21-APR-1999; 9905-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154029.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

OS Zea mays subsp. mays.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126268.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132563.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151308.
PR 30-AUG-1999; 99US-0151309.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155146.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159295.
PR 13-OCT-1999; 9905-0159285.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.
```

Query Match 11.8%; Score 48.8; DB 21; Length 486;
Best Local Similarity 62.1%; Pred. No. 3.2e-06;
Matches 77; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
OY 25 gggaagaataatacagaatcaagaatggtacacggggttctatatacaaccgatg 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 gagcaccatcatgagggagctgcgtctctccgtgtgtctatatacaccacgacc 245
OY 85 ttctgtcgtgaggttttgaccacacacgtttaacctgtatcaatacaacatgccaa 144
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 tgcgtgtcgtgtgattgattcagtcagcaggtgtccctgcgtatcaatacattgattccca 305
OY 145 tcaa 148
DB 306 ccca 309
```

RESULT 4
ID C43533
CA3533 standard; DNA; 496 BP.
XX C43533;
XX
DT 17-OCT-2000 (first entry)
XX
DE zea mays DNA fragment SEQ ID NO: 39582.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
pathway; promoter; termination sequence; corn; ss.

```
XX OS zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
```


XX	antInflammatory; gene therapy; diagnosis; ss.
XX	
OS	Homo sapiens.
XX	
XX	Key
XX	Location/Qualifiers
XX	107..1546
XX	CDS
XX	/*tag= a
XX	
XX	MO9964596-A2.
XX	
XX	16-DEC-1999.
XX	
XX	11-JUN-1999; 99WO-US13281.
XX	
XX	12-JUN-1998; 98US-0089029.
XX	PR 29-JUL-1998; 98US-0094575.
XX	PR 14-OCT-1998; 98US-0104624.
XX	
XX	(INCYTE) INCYTE PHARM INC.
XX	
XX	Lal P, Yue H, Tang YF, Hillman JL, Bandman O, Corley NC;
XX	Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DAM;
XX	WPI: 2000-116543/10.
XX	P-PSDB: Y58616.
XX	
XX	Claim 9; Page 126-127; 150pp; English.
XX	
XX	The present sequence is that of Incyte clone 2683322 encoding new
XX	human protein regulating gene expression PRGE-9 (see Y58616). The
XX	CDNA was initially isolated from ileum CDNA library SINICT01,
XX	CC and the full-length sequence assembled from overlapping clones from a
XX	number of libraries. PRGE-9 is expressed in reproductive, CC
XX	gastrointestinal and cardiovascular tissues associated with foetal,
XX	CC cancer and inflammation diseases, disorders or conditions. The
XX	protein product is characterised as an RNA helixase and DEAD-box
XX	CC subfamily protein. The invention provides PRGE polypeptides (see
XX	CC Y58608-38) and polynucleotides (see 257839-69), expression vectors,
XX	CC host cells, antibodies, agonists and antagonists. It also provides
XX	CC methods for diagnosing, treating or preventing disorders associated
XX	CC with expression of PRGE. Polynucleotides are also used as sources
XX	CC of probes and primers for diagnosis and monitoring of disease, also
XX	CC for detecting related sequences and in gene mapping.
XX	
S0	Sequence 1797 BP; 511 A; 457 C; 469 G; 360 T; 0 other;
	Query Match 11.4%; Score 47.4; DB 21; Length 1797;
	Best Local Similarity 50.5%; Pred. No. 1.7e-05;
	Matches 145; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
Db	14 ggaagcgtggtgggggaaagaataacacaggaattcaagaatggtctacaccgggtcttat 73
Db	1210 ggtggaagcagagggcgctgagatgagtcgagtcgagcgaagggcagaagagaaagtgtt 1269
Db	74 atcaacagatcttctgctgcaggtgttgaccacaagcacaggttaacctgtacataacta 133
Db	1270 gaccacccaactgtgtgcccgcgcgcatgagtgttgaaacaagtgctgtcgttacaact 1329
Db	134 cgaacatgccaatcaaatctgtgtacagaagatgaacctgatatgagtggttactgcag 193
Db	1330 tgatcttcocgtga-----caaggaacgggaatactctgacatgtgactactgcagc 1383
Db	194 aattggcagagctggtggccttcttgcccgaaaggtgtctgttcaactgtgtgtgtca 253
Db	1384 gatcggcgccacggcgctgttggaagaagggtccttgacgtgacacatgltgacagcaa 1443
Db	254 aaccattatgtgtgatgaaagaagaatgaagaactatttcacggaca 300
Db	1444 gcaacgatgaacatctctgaaacagaatccacgagacatttataaga 1490

```

RESULT      6
ID           F13835
ID           F13835 standard; cDNA; 1205 BP.
XX
XX           F13835;
XX
XX           13-MAR-2001 (first entry)
XX
XX           Aspergillus oryzae EST SEQ ID NO:6358.
XX
XX           Multiple gene expression; filamentous fungal cell; EST;
XX           expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX           Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX           culture condition; environmental stress; spore morphogenesis;
XX           metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX           Aspergillus oryzae.
XX
XX           WO2000056762-A2.
XX
XX           28-SEP-2000.
XX
XX           22-MAR-2000; 2000WO-US07781.
XX
XX           22-MAR-1999; 99US-0273623.
XX
XX           (NOVO ) NOVO NORDISK BIOTECH INC.
XX           (NOVO ) NOVO NORDISK AS.
XX
XX           Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IC, Olsen PB;
XX           WPI: 2000-594572/56.
XX
XX           Monitoring differential expression of genes in filamentous fungal cells
XX           uses fluorescence-labeled nucleic acids isolated from the cells and a
XX           substrate of expressed sequence tags -
XX
XX           Claim 88; Page 2611-2612; 3161pp; English.
XX
XX           The present invention describes a method for monitoring differential
XX           expression of genes in a first filamentous fungal (FF) cell relative to
XX           expression of the same genes in one or more second filamentous fungal
XX           cells. The method uses fluorescence-labeled nucleic acids isolated from
XX           the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX           are used in the methods for monitoring differential expression of genes
XX           in a first filamentous fungal (FF) cell relative to expression of the
XX           same genes in one or more second filamentous fungal cells. Monitoring
XX           the global expression of genes from FF cells allows the production
XX           potential of the microorganisms to be improved. New genes may be
XX           discovered, possible functions of unknown open reading frames can be
XX           identified and gene copy number variation and stability can be
XX           monitored. The expression of genes can be used to study how FF cells
XX           adapt to changes in culture conditions, environmental stress, spore
XX           morphogenesis, recombination, metabolic or catabolic pathway engineering.
XX           using ESTs provides several advantages over genomic or random cDNA
XX           clones including elimination of redundancy as one spot on an array
XX           equals one gene or open reading frame, and organisation of the
XX           microarrays based on function of the gene products to facilitate
XX           analysis of the results. F07478 to F1247 represents ESTs from Fusarium
XX           venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
XX           F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
XX           F15337 represents ESTs from Trichoderma reesei, which are all
XX           specifically claimed in the present invention.
XX
XX           Sequence 1205 BP; 243 A; 336 C; 296 G; 327 T; 3 other;
XX
XX           Query Match      11.4%; Score 47; DB 21; Length 1205;
XX           Best Local Similarity 58.0%; Pred. No. 1.9e-05;
XX           Matches      83; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

QY 11 ggcgcgcgcgtggtggggaataataacagggaattcaagaatgtgctacacggggttct 70
DB 500 ggagcaggttcacgtcgtgattcatcgaaggatcgcctcgtgtctctctcgtgtcct 559
QY 71 tatataaccagatgtctctcgcaggttttcaccaagcacagcttaacctgtcatcaa 130
DB 560 gatcgcacattgcactcttcttgcccgctgtatcgcagtcacagcaggtctccctgtcatcaa 619
QY 131 ctacgacatgcattcaaatcttg 153
DB 620 ctacgacatcctcctgcacacgcgtg 642

RESULT 7
C34215
ID C34215 standard; DNA; 1621 BP.
XX
AC C34215;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5867.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX
PN EPI03405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

Db 1091 ggagcatcatcatgagagagtcaggtcgtcttcctccggtgttcctcatccaccctgacc 1150
Qy 85 tctctgtcgtcaggttttgaccacgacacgttaaccttgcatacctaagacatgccaa 144
Db 1151 tcttagctcgtggtatgatagtccaacaagtcctccctggtatacctaactgatactcccaa 1210

RESULT 9
C47750
ID C47750 standard; DNA; 1763 BP.
AC C47750;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 54978.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic;
KM pathway; promoter; termination sequence; corn; ss.
XX
XX Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 22-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130691.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.0%; Score 45.4; DB 21; Length 1763;
Best Local Similarity 61.3%; Pred. No. 7.6e-05;
Matches 73; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 40 aggaattcaagaatgctacacggggttctatatacaacggatttctgctcgaggtt 99
Db 1166 agaattcaagaagaagacaaagagattctgtgctactgatttggttgtaggggaa 1225
Qy 100 ttgaccagaacacgttaacctgtcatcaactacgaatcgcaatcaatttggttaca 158

Db 1226 tgcacatgcaacgtgtcaacatgtgcgtaactatgatatgctatccgtgataca 1284
RESULT 10
ID C38963 standard; DNA; 1510 BP.
AC C38963;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 22844.
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22844.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN Epi1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139464.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.

PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158332.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 25-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match 10.9%; Score 45; DB 21; Length 1510;
Best Local Similarity 60.0%; Pred. NO. 9.6e-05;
Matches 75; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

OY 145 tcaaa 149
| |
Db 1167 cccaa 1171

RESULT 11
ID C40626 standard; DNA: 737 BP.
C40626;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 28988.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136332.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139761.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142809.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.88; Score 44.6; DB 21; Length 737;
Best Local Similarity 65.7%; Pred. No. 9.3e-05;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 45 ttcaaatggtctaccacgggtctctatatacaacgagtgcttgcaggtttgac 104
DB 180 ttcaaggaaggacacagaagatccttctgtgcgactgacttgtaagaaggaggttgcac 239
QY 105 caagcacagcttaacctgtcatcaactacacatgcaca 143
DB 240 atagagcgtgcacacattgttatcaactatgacatgcaca 278

RESULT 12
C38774
ID C38774 standard; DNA; 1082 BP.
XX
AC C38774;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22184.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135263.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

C34340
ID C34340 standard: DNA; 1636 BP.
XX
AC C34340;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6297.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0148368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144684.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.

PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 30-AUG-1999: 99US-0151080.
PR 31-AUG-1999: 99US-0151303.
PR 01-SEP-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160880.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 10 6%; Score 44; DB 21; Length 1591;
Best Local Similarity 59.7%; Pred. No. 0.00021;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 25 gggaagaataacgaatcaagaatgctacacgggtctctataccaacgaty 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1036 gggatgcatatagggggaattcaggtcgcgcgaactctggtgctaataccaagaty 1095
QY 85 ttcttgctgaggttttgaccaagacaagcttaacctgtcatccaactacgacatgcaa 144
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1096 ttggctcgagacgtgacgtccagcaggtttcaactgtcatcaataatlatgatcccaa 1155
QY 145 tcaa 148
||
Db 1156 ataa 1159

RESULT 15
C77791
ID C77791 standard; cDNA; 1987 BP.

XX C77791;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:185.
 XX
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
 KW antidiabetic; antiaesthetic; antineumatic; antiaesthetic; antiviral;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;
 KW vasotropic; antiparasitic; antineoplastic; gene therapy; inflammation;
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20005350-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05882.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587533/55.
 DR P-PSDB: B43582.
 XX
 PT Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX
 PS Claim 1; Page 762; 2352pp; English.
 XX
 CC C77607 to C78448 encode the human cancer associated proteins given in
 CC B43398 to B44239. The proteins can have activities based on the tissues
 CC and cells the genes are expressed in. Example of activities include:
 CC cytostatic; proliferative; vulnerable; immunomodulator; antidiabetic;
 CC antiaesthetic; antineumatic; antiaesthetic; antineoplastic;
 CC antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 CC antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 CC neuroprotective; cardiac; thrombolytic; coagulant; nocotropic;
 CC vasotropic; antiparasitic and antineoplastic. The polynucleotides and
 CC polypeptides can be used for preventing, treating or ameliorating medical
 CC conditions and diagnosing pathological conditions. Polynucleotides,
 CC polypeptides, antibodies, agonists and antagonists from the present
 CC invention may be used to treat immune disorders by activating or
 CC inhibiting the proliferation, differentiation or mobilisation of immune
 CC cells, to treat disorders of haematopoietic cells, autoimmune disorders,
 CC allergic reactions, graft versus host disease and organ rejection,
 CC modulate haemostatic or thrombolytic activity, modulate inflammation,
 CC cancers, cardiovascular disorders, neurological disease and bacterial or
 CC viral infections. The peptides, nucleotides, antibodies, agonists and
 CC antagonists may be also be used in drug screens. C78449 to C78457 and
 CC B44240 represent sequences used in the exemplification of the present
 CC invention.
 XX
 QX Sequence 1987 BP; 689 A; 381 C; 432 G; 483 T; 2 other;
 QY

Oy 83 tgccttgcctcgaagctttttacaaagcacacscgttaaccttgcatacctaagaatgcc 142
 Db 1444 ctctcctgcagaagggatctgattttaaagtgctgcaccttgcatacctaagcttccc 1493
 Oy 143 aa 144
 Db 1494 aa 1495

Search completed: May 8, 2001, 15:14:01
Job time: 43382 sec

Query Match	10.58	Score 43.6	DB 21	Length 1987
Best Local Similarly	59.88	Pred. No. 0.00031		
Matches 73	Conservative 0	Mismatches 49	Indels 0	Gaps 0

Oy 23 gggggaataataacaggaaatccaagatgtcacccgggtcttatatacaaccga 82
| | | | | | | | | | | | | | | |
Db 1374 gaagaataacacagtcacagtttceagacggaaaactctggtcttgatttgtacgc 1433

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:15 ; Search time 132.33 Seconds
(without alignments)
546.252 Million cell updates/sec

Title: US-09-619-643-4

Perfect score: 414

Sequence: 1 cggacgcgtggcgagcg.....gtgtttgtcgtacctctgt 414

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PCFUS.COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	10.3	1245	4 US-09-318-443-1	Sequence 1, Appl
2	41.2	10.0	1536	4 US-09-318-443-5	Sequence 5, Appl
3	41.2	10.0	1682	4 US-09-318-443-7	Sequence 7, Appl
4	40.6	9.8	1618	2 US-08-533-669A-9	Sequence 9, Appl
5	40.6	9.8	1618	2 US-08-607-509-1	Sequence 1, Appl
6	40.6	9.8	1618	2 US-08-454-036-1	Sequence 1, Appl
7	40.6	9.8	1618	2 US-08-634-642-1	Sequence 1, Appl
8	40.6	9.8	1618	2 US-08-989-370-1	Sequence 1, Appl
9	40.6	9.8	1618	5 PCT-US95-05064-1	Sequence 1, Appl
10	40.6	9.8	1867	2 US-08-607-509-3	Sequence 3, Appl
11	40.6	9.8	1867	2 US-08-634-642-3	Sequence 3, Appl
12	40.6	9.8	1867	3 US-08-989-370-3	Sequence 3, Appl
13	39.8	9.6	235	2 US-08-245-511-19	Sequence 19, Appl
14	39.8	9.6	235	2 US-08-600-993A-19	Sequence 19, Appl
15	37.4	9.0	3825	4 US-09-208-742-3	Sequence 3, Appl
16	37.4	9.0	43676	4 US-09-356-952-12	Sequence 12, Appl
17	36	8.7	2061	4 US-09-008-271A-16	Sequence 16, Appl
18	34	8.2	1446	1 US-08-586-024-5	Sequence 5, Appl
19	32.6	7.6	1254	5 PCT-US96-05320A-894	Sequence 894, Appl
20	31.4	7.6	1254	4 US-09-058-489-14	Sequence 14, Appl
21	31.4	7.5	5322	4 US-09-058-489-13	Sequence 13, Appl
22	31.2	7.5	644	3 US-08-911-894-71	Sequence 71, Appl
23	31.2	7.5	1347	2 US-08-959-749-1	Sequence 1, Appl
24	30.8	7.4	2934	4 US-09-149-934-2	Sequence 2, Appl
25	30.2	7.3	3742	1 US-08-694-915-5	Sequence 5, Appl
26	29.8	7.2	2319	4 US-09-058-489-90	Sequence 90, Appl
27	29.8	7.2	4416	4 US-09-058-489-17	Sequence 17, Appl

C	28	29.8	7.2	7218	1 US-08-232-463-14	Sequence 14, Appl
C	29	29.6	7.1	1521	3 US-08-931-952-3	Sequence 3, Appl
C	30	29.6	7.1	1521	3 US-08-272-247-3	Sequence 3, Appl
C	31	29.6	7.1	1521	5 PCT-US95-08560-3	Sequence 3, Appl
C	32	29.2	7.1	1063	1 US-08-325-562-1	Sequence 1, Appl
C	33	29.2	7.1	1063	1 US-08-437-795-1	Sequence 1, Appl
C	34	29.2	7.1	5319	1 US-08-169-927-1	Sequence 1, Appl
C	35	28.8	7.0	8298	5 PCT-US93-03076-1	Sequence 1, Appl
C	36	28.6	6.9	3728	1 US-08-111-939-1	Sequence 1, Appl
C	37	28.4	6.9	1350	6 5468481-2	Patent No. 5468481
C	38	28	6.8	3981	2 US-08-955-138-2	Sequence 2, Appl
C	39	27.8	6.7	3805	4 US-09-513-729B-10	Sequence 10, Appl
C	40	27.6	6.7	1800	4 US-09-039-773A-1	Sequence 1, Appl
C	41	27.4	6.6	1350	4 US-08-462-351-1	Sequence 1, Appl
C	42	27.4	6.6	1619	2 US-08-991-946A-2	Sequence 2, Appl
C	43	27.4	6.6	2881	2 US-08-570-227A-1	Sequence 1, Appl
C	44	27.4	6.6	2881	4 US-09-077-991-1	Sequence 1, Appl
C	45	27.4	6.6	17949	4 US-09-087-465-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-09-318-443-1
; Sequence 1, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briavanlou, Ali
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Xenopus laevis
US-09-318-443-1

Query Match      10.3%; Score 42.8; DB 4; Length 1245;
Best Local Similarity 58.7%; Pred. NO. 9.8e-05;
Matches 74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 23 gggggaagaataatacaagaattcaagatgctacacgggggtttcttatcaacga 82
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 954 gaggagatcatcatgaagaattcgcgtgtgcaagccgagctcatatcaacga 1013
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 83 tgtcttgcgcgaagtttgcaacgacacgttaacctgttcatacaactacacgac 142
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1014 cgttggccgcgaagattgtgacacagtgctctcttattatcaactatgctcttc 1073
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 143 aatcaa 148
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1074 caataa 1079

RESULT 2
US-09-318-443-5
; Sequence 5, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briavanlou, Ali
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
```



```

: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.404C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206)682-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1618 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 115..1326
: OS-08-607-509-1

```

	Query Match	9.8%	Score 40.6:	DB 2:	Length 1618;
	Best Local Similarity	64.2%:	Pred. No. 0.00059;		
	Matches 61; Conservative	0;	Mismatches 34;	Indels	0; Gaps
Oy	54	ggctacaccgsggttttcattatacaacgcagatgttcttgctgaggtttgaccaagcacac	113		
Db	1060	GGCAGCTCCCGTCGTCTGTAACGCACCACCTCGTGGGCCGGGCATGACGTGCACCAC	1119		
Oy	114	gttaacctgtcatcactaacctacagctgaatcaa	148		
Db	1120	GTTGAACATGTCATCATCAACTTGCACTTGCACCTGGCGACGAA	1154		

```

1      RESULT      6
2      US-08-454-036-1
3      : Sequence 1, Application US/08454036
4      : Patent No. 5876966
5      : GENERAL INFORMATION:
6      : APPLICANT: Reed, Steven G.
7      : TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEMENT OF
8      : NUMBER OF SEQUENCES: 2
9      : CORRESPONDENCE ADDRESS:
10     : ADDRESSEE: SEED and BERRY
11     : STREET: 6300 Columbia Center, 701 Fifth Avenue
12     : City: Seattle
13     : STATE: Washington
14     : COUNTRY: USA
15     : ZIP: 98104-7092
16     : COMPUTER READABLE FORM:
17     : MEDIUM TYPE: Floppy disk
18     : COMPUTER: IBM PC compatible
19     : OPERATING SYSTEM: PC-DOS/MS-DOS
20     : SOFTWARE: PatentIn Release #1.0, Version #1.30
21     : CURRENT APPLICATION DATA:
22     : APPLICATION NUMBER: US/08/454,036
23     : FILING DATE: 30-MAY-1995
24     : CLASSIFICATION: 514
25     : ATTORNEY/AGENT INFORMATION:
26     : NAME: Kadlecsek, Ann T.
27     : REGISTRATION NUMBER: P-39,244
28     : REFERENCE/DOCKET NUMBER: 210121.404C1
29     : TELECOMMUNICATION INFORMATION:
30     : TELEPHONE: (206)682-6031
31     : TELEFAX: (206)682-6031
32     : TELEEX: 3723836
33     : INFORMATION FOR SEQ. ID NO: 1:
34     : SEQUENCE CHARACTERISTICS:
35     : LENGTH: 1618 base pairs
36     : TYPE: nucleic acid
37     : STRANDEDNESS: single
38     : TOPOLOGY: linear
39     : MOLECULE TYPE: cDNA
40     : NAME/KEY: CDS
41     :
42     :
43     :
44     :
45     :
46     :
47     :
48     :
49     :
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :

```

LOCATION: 115..1326
US-08-454-036-1

Query Match	9.8%;	Score 40.6;	DB 2;	Length 1618;
Best Local Similarity	64.2%;	Pred. No. 0.00059;		
Matches 61; Conservative	0;	Mismatches 34;	Indels 0;	Gaps 0;

Accession	Sequence	Position
Oy	ggctcaacccgggggtctctatatacaacggatggtctctgcgcgggtttbaccagaacac	113
Db	ggcagctccgcggcggcgtcgtatgacacgcgacctgtggccggcgatcgactgcacac	1119
Oy	gttaacctgtgacatcaactcgcgatgccaatcaa	148
Db	gtgaaacatcgtcatcaactctggccttcggacgcaa	1154

```

RESULT 7
US-08-634-642-1
: Sequence 1, Application US/08634642
: Patent No. 5879687
:
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF
: TITLE OF INVENTION: PROTECTIVE IMMUNE RESPONSES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/634,642
: FILING DATE: 18-Apr-1996
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.404C4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206)682-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1618 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 115..1326
:
US-08-634-642-1

```

```
RESULT 8
US-08-989-370-1
; Sequence 1, Application US/08989370
; Patent No. 6013268
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,370
; FILING DATE: 12-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.404C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..1326
; US-08-989-370-1

Query Match          9.8%; Score 40.6; DB 3; Length 1618;
Best Local Similarity 64.2%; Pred. No. 0.00059;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 54 ggctaacccggggtctctatcaaccgagtgtctgtcgcgagggtttgaccaagcacac 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1060 ggcagctctccggcgtctgctgaacgacgacctgtggtgcccgcgcatgcagctgcacacac 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 114 gttacctgtcatcaactacgaacatgcacatcaa 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1120 gtgaacatgctcatcaacttcgacgtgccgacgaa 1154

RESULT 9
PCT-US95-05064-1
; Sequence 1, Application PC/TUS9505064
; GENERAL INFORMATION:
; APPLICANT: Corixa, Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEMENT OF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05064
; FILING DATE: 24-APR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Radlecek, Ann T.
; REGISTRATION NUMBER: P-39,244
; REFERENCE/DOCKET NUMBER: 210121.404PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..1326
; PCT-US95-05064-1

Query Match          9.8%; Score 40.6; DB 5; Length 1618;
Best Local Similarity 64.2%; Pred. No. 0.00059;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 54 ggctaacccggggtctctatcaaccgagtgtctgtcgcgagggtttgaccaagcacac 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1060 ggcagctctccggcgtctgctgaacgacgacctgtggtgcccgcgcatgcagctgcacacac 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 114 gttacctgtcatcaactacgaacatgcacatcaa 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1120 gtgaacatgctcatcaacttcgacgtgccgacgaa 1154

RESULT 10
US-08-607-509-3
; Sequence 3, Application US/08607509
; Patent No. 5876735
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/607,509
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.404C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
```



```

: TITLE OF INVENTION: ACCELLULAR VACCINES BASED THERON
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/245,511
: FILING DATE: 18-MAY-1994
: CLASSIFICATION: 424
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/116,541
: FILING DATE: 01-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-069 CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 235 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: R6
: IMMEDIATE SOURCE:
: CLONE: SPRU17
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..234
: US-08-245-511-19

Query Match          9.6%; Score 39.8; DB 2; Length 235;
Best Local Similarity 60.7%; Pred. No. 0.00046;
Matches 65; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 37 tacagaatcaagaatggctacacccgggtcttatatcaacgaltgtctgtcgag 96
DB 122 TTCGTACTTTAAATAATGCAATCTTGATGTTTGGTTCGACAGACGTTGCAGCCGCGTg 181

QY 97 gtttgaaccaagacacgttaacctgtcatcaactaagacatgcca 143
DB 182 GTTTGATATTTCAGGTGTGACCCATGCTACACACTACATATTCCA 228

RESULT 14
US-08-600-993A-19
: Sequence 19, Application US/08600993A
: Patent No. 5981229
: GENERAL INFORMATION:
: APPLICANT: Masure, H Robert
: APPLICANT: Pearce, Barbara J
: APPLICANT: Tuomanen, Elaine
: TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
: TITLE OF INVENTION: ACCELLULAR VACCINES BASED THERON
: NUMBER OF SEQUENCES: 59
```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Klauber & Jackson
: STREET: 411 Hackensack Avenue
: CITY: Hackensack
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/600,993A
: FILING DATE: 1-MAR-1996
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/245,511
: FILING DATE: 18-MAY-1994
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/116,541
: FILING DATE: 01-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Jackson Esq., David A.
: REGISTRATION NUMBER: 26,742
: REFERENCE/DOCKET NUMBER: 600-1-069 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201 487-5800
: TELEFAX: 201 343-1684
: TELEX: 133521
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 235 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus pneumoniae
: STRAIN: R6
: IMMEDIATE SOURCE:
: CLONE: SPRU17
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..234
: US-08-600-993A-19

Query Match          9.6%; Score 39.8; DB 2; Length 235;
Best Local Similarity 60.7%; Pred. No. 0.00046;
Matches 65; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 37 tacagaatcaagaatggctacacccgggtcttatatcaacgaltgtctgtcgag 96
DB 122 TTCGTACTTTAAATAATGCAATCTTGATGTTTGGTTCGACAGACGTTGCAGCCGCGTg 181

QY 97 gtttgaaccaagacacgttaacctgtcatcaactaagacatgcca 143
DB 182 GTTTGATATTTCAGGTGTGACCCATGCTACACACTACATATTCCA 228

RESULT 15
US-09-208-742-3
: Sequence 3, Application US/09208742
: Patent No. 6174679
: GENERAL INFORMATION:
: APPLICANT: Kaufmann, Joerg
: TITLE OF INVENTION: CIP150/htraf150 is Necessary for Cell
: TITLE OF INVENTION: Cycle Progression
: FILE REFERENCE: 1453.002
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:30 ; Search time 3095.41 Seconds
(without alignments)
1168.422 Million cell updates/sec

Title: US-09-619-643-4

Perfect score: 414
Sequence: 1 cggacgcgtggcgagcg.....gtgttttgcgtgacctctgt 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*
48: gb_est48:*
49: gb_est49:*
50: gb_est50:*
51: gb_est51:*
52: gb_est52:*
53: gb_est53:*
54: gb_est54:*
55: gb_est55:*
56: gb_est56:*
57: gb_est57:*
58: gb_est58:*
59: gb_est59:*
60: gb_est60:*
61: gb_est61:*
62: gb_est62:*
63: gb_est63:*
64: gb_est64:*
65: gb_est65:*
66: gb_est66:*
67: gb_est67:*
68: gb_est68:*
69: gb_est69:*
70: gb_est70:*
71: gb_est71:*
72: gb_est72:*
73: gb_est73:*
74: gb_est74:*
75: gb_est75:*
76: gb_est76:*
77: gb_est77:*
78: gb_est78:*
79: gb_est79:*
80: gb_est80:*
81: gb_est81:*
82: gb_est82:*
83: gb_est83:*
84: gb_est84:*
85: gb_est85:*
86: gb_est86:*
87: gb_est87:*
88: gb_est88:*
89: gb_est89:*
90: gb_est90:*
91: gb_est91:*
92: gb_est92:*
93: gb_est93:*
94: gb_est94:*
95: gb_est95:*
96: gb_est96:*
97: gb_est97:*
98: gb_est98:*
99: gb_est99:*
100: gb_est100:*
101: gb_est101:*
102: gb_est102:*
103: gb_est103:*
104: gb_est104:*
105: gb_est105:*
106: gb_est106:*
107: gb_est107:*
108: gb_est108:*
109: gb_est109:*
110: gb_est110:*
111: gb_est111:*
112: gb_est112:*
113: gb_est113:*
114: gb_est114:*
115: gb_est115:*
116: gb_est116:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	221.6	53.5	539	166	BE400232
2	209.4	50.6	641	116	AW497526
3	206.4	49.9	597	23	AW670627
4	167.2	40.4	471	143	BF070035
5	167.2	40.4	688	114	AW348529
6	165.8	40.0	773	114	AW349427
7	163.8	39.6	521	111	AW153139
8	163.8	39.6	549	138	BE658854
9	154.8	37.4	518	163	BE131561
10	154.8	37.4	598	21	AW486996
11	153	37.0	484	173	BC101919
12	147.8	35.7	459	121	AW933319
13	147.8	35.7	578	167	BE437127
14	144.6	34.9	647	24	AW782217
15	138.8	33.5	593	24	AW777083
16	122.4	29.6	419	148	BF425946
17	120.6	29.1	720	112	AW220283
18	99.8	24.1	615	119	AW760007

BASE COUNT 216 a 166 c 106 g 153 t
ORIGIN

Query Match 50.6%; Score 209.4; DB 116; Length 641;
Best Local Similarity 94.8%; Pred. No. 6.3e-53;
Matches 238; Conservative 0; Mismatches 11; Indels 2; Gaps 2;

Oy 152 tggtaacaagaatgaacacgataatgaaggctgactgtcacagaatgtgcagagctgagc 211
|||||
Db 641 TGGTACAAAGATGACACTGATTT-TGAGGTGACTTGC-CAGAAATGGCAGAGCTGGGG 584
Oy 212 ctttgccggaagagtgctgctgttcaactgtgtgtgtgtcaaacgacatagtgtgat 271
|||||
Db 583 CTTTGGCCGGAAGGCTGTGTCAACTGTGTGTCAAAACCGATTAATGTTGTGAT 524
Oy 272 gaagaagaatcgagaactatttcacagacagtgctgacctgtgttcggaattgcagaactga 331
|||||
Db 523 GAAGAAAGATCGAGACTATTTCACAGCACAGTGTGCTTGAGTTCCGAAATTTGCCAACGCTGA 464
Oy 332 aaaaataatgaagctgacatcacaagatgcagcttactcttaagttaaccaggaact 391
|||||
Db 463 AGAAGATTTTGAGGCTGCTCAAGATGACAGTTTACTTTAAAGTAACACGAAACTT 404
Oy 392 gtggtgtttt 402
|||||
Db 403 GTGCTGTGTTT 393

RESULT 3
LOCUS A1670627 597 bp mRNA EST 02-FEB-2000
DEFINITION 605035H12.X1 605 - Endosperm cDNA library from Schmidt lab Zea mays
CDNA, mRNA sequence.

ACCESSION A1670627
VERSION A1670627.1 GI:4837537
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
; Andropogoneae; Zea.
1 (bases 1 to 597)

REFERENCE 1 Walbot V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605035 row: H column: 12.

FEATURES
source 1..597
Location/Qualifiers
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH3(alpha)"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; site_1: EcoRI;
site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT 204 a 158 c 101 g 134 t
ORIGIN

Query Match 49.9%; Score 206.4; DB 23; Length 597;
Best Local Similarity 95.1%; Pred. No. 5.1e-52;

Matches 213; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 179 ggttactgtcacagaattgtgcagagctgtgagccttgccggaagtgctgtgtca 238
|||||
Db 597 GGTGTACTTGCAAGAAATTTGGCAGAGCTGGGGCTTTGGCCGGAAGGCTGTGTCA 538
Oy 239 cttgtctgtgtgtcaaacgacatagtgtgtgtgaagaagatcgaaactattccagca 298
|||||
Db 537 CTTGCTGTGTGTCAAAACCGATTAATGTTGTGATGAAGAAGATCGAGACTATTTCACGA 478
Oy 299 cagttgtccttattgttcgaattgcagaactgtgaataatatttgaggtcgtactaaga 358
|||||
Db 477 CAGTGTGCTGAGGTTGGGAATTTGCCAACGTGAAGAAGATTTTGAAGGCTGCTCAAGA 418
Oy 359 tgcagttacttacttaagttaaccaggaactgtgtgtttt 402
|||||
Db 417 TCGAGTTTACTTTAAAGTAACACGGAACACTTGTGTGTTT 374

RESULT 4
LOCUS BF070035 471 bp mRNA EST 17-OCT-2000
DEFINITION st62h09.y1 Gm-cl053 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl053-1506.5' similar to TR:Q9SCQ0 Q9SCQ0 RNA HELICASE-LIKE
PROTEIN, ; mRNA sequence.
ACCESSION BF070035
VERSION BF070035.1 GI:10847330
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 471)

REFERENCE 1 Shoemaker R., Kelm P., Vodkin L., Erpeland J., Coryell V., Khanna
A., Bolla B., Merritt M., Hillier L., Kucaba T., Martin J., Beck C.,
Wylie T., Underwood K., Steptoe M., Theising B., Allen M., Bowers
Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurr
R., Riller E., Kohn S., Shih T., Jackson Y., Cardenas M., McCann
R., Waterston R. and Wilson R.
TITLE Public Soybean EST Project
UNPUBLISHED (1999)
JOURNAL
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

FEATURES
source 1..471
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl053-1506"
/clone_lib="Gm-cl053"
/tissue_type="whole seedling, 3 week old, greenhouse
grown"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from whole seedlings of 3
week old greenhouse grown plants. Complementary DNA was
synthesized from mRNA using a primer consisting of a 3'
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adapters were ligated to the blunt-ended

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 773)
 Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Correll, V., Erpelting, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L., and Lewin, H.,
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Other_ESTS: A1460952
 Contact: Vodka, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics

University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodka@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers

FEATURES
 SOURCE
 1..773
 /organism="Glycine max"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone="Gm-r1021-2516"
 /clone_lib="Gm-r1021"
 /tissue_type="root"
 /lab_host="XLI0-Gold"
 /note="Vector: pBluescript II XR; Site: 1: EcoRI; Site: 2: XhoI; Library Gm-r1021 is a sequence-driven, reacked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. StrataGene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Correll, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@na.u.edu, virginia.correll@na.u.edu. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html. Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/Diotech/keck.html."
 BASE COUNT 235 a 182 c 116 g 219 t 21 others
 ORIGIN

Query Match 40.0%; Score 165.8; DB 114; Length 773;
 Best Local Similarity 69.0%; Pred. No. 1.2e-39;
 Matches 240; Conservative 0; Mismatches 105; Indels 3; Gaps 1;
 23 gggggaagaataatcaagaatcaaaatgctacaccgggtcttatatacaccga 82
 669 GAGANNNAAGTTGTCAAGAGAGTCAAAAGATGTTTACCCCAAGTCTTATATACACACA 610
 83 tgtctgtcagaggtttgacaaagcacacgttaacctgtcaactacagacatgc 142
 609 TATTCTTGTCTGCTTGTTCATCAGACACAGCTTATTGGTCATCATTAATATGATCTTCC 550
 143 aatcaaatgtgtacaaagatgaacctgatatagagtgctacttcacagaattgcag 202
 549 CAAGAAATACGGTGTACGTGATGATGACCATATATGAAGTGTATTTCACACAGGTTGCAG 490

Qy 203 agctggcgcttgcgcggaaggtcgtgttcaactgtcgtgtgtgtcaaacgata 262
 489 AGCTGGCGCTTTTGGCGCAGAGGGCTCTATTAACTGATATGTGTGATTTGGATGA 430
 Qy 263 tgtgtgtaagaagaatgaagactatccagacagtggtgcctgtatgttcgaatg 322
 429 AAGGCTCATGTGCAAGATGACAGACCATTTTGGCACTGCTGTAAGTCAAGTTCG--ACG 373
 Db 372 ACAAACTGTGAAGATATTAAGCTCTCTCAAGGAACGGGTTACT 325

RESULT 7
 AM153139 521 bp mRNA EST 17-JUL-2000
 LOCUS se36e02.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl015-3315 5' similar to WP:F57B9.6 CE01341 INF-1: EIF-4A ;
 DEFINITION
 mRNA sequence.
 AM153139
 VERSION AM153139.1 GI:6201039
 KEYWORDS
 SOURCE soybean.
 ORGANISM Glycine max

REFERENCE
 AUTHORS
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 521)
 Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Correll, V., Khanna, A., Bolla, B., Marra, P., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pepe, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 CONTACT: Shoemaker R/Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Insert length: 637 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 423.
 Location/Qualifiers

FEATURES
 SOURCE
 1..521
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-3315"
 /clone_lib="Gm-cl015"
 /tissue_type="Mature flowers, field grown plants"
 /lab_host="XLI0-Gold"
 /note="Vector: pBluescript II XR; Site: 1: EcoRI; Site: 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XLI0-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

```

BASE COUNT      150 a      88 c      120 g      162 t      1 others
ORIGIN

Query Match
Best Local Similarity 39.6%; Score 163.8; DB 111; Length 521;
Matches 244; Conservative 0; Mismatches 112; Indels 3; Gaps 1;

OY 23 gggggaagaataatacagaatcgaagatgctacacgggtttatatacaccga 82
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 32 GAGAGACAAAGTTGTCAGAGAGTTCAAGATGTTGATCAAGTTTATATCAACAGA 91
OY 83 tttcttcgcgcgaaggtttgacacacacgttaacctgttatacacaacacagcc 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 92 TATTTCTTCTCGTGGCTTTTATATCAACAGTTAATTTGTCATCAATTAATCTTCC 151
OY 143 aatcaaatctgtacagaagatgaacctgtatctgaaggttacttgcacagaattggcag 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 152 AAATTAACACAGTTTACGGGATGACACAGATTATGAGTGTATGTCACAGGGTTGGCAG 211
OY 203 agctggcgcttggccggaaggtgctgtgtcaactgtgtgtgtgtaaacacgatta 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 212 AGCTGGCGCTTTGGCGCAAGGGCGCTGTATTAACTGATATGATGAATGATGA 271
OY 263 tgtgtgatgaagaagatcgaaagactatctccacagatgtgcctatgttgcgaattg 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 272 AAGGCTCATGTCGAAGATTGAACCAATTTGGCATTTGTGTAAGTGAAGTGCAGCT-- 329
OY 323 gcaacgtgaataataataatagagctgatacgaagatgcacgttactttaaagttacc 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 330 -CAAAGTGTGAAGATATTAAGCCGCTCTCAAGAGACGGGTTACTGCAATGATGCC 387

RESULT 8
BE658854 549 bp mRNA EST 06-SEP-2000
LOCUS BE658854 GM700007B10H10 GM-r1070 Glycine max cDNA clone GM-r1070-2683 3'
DEFINITION mRNA sequence.
ACCESSION BE658854
VERSION BE658854.1 GI:9984753
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max.
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
    Fabales; Fabaceae; Papilionoideae; Glycine.
    1 (bases 1 to 549)
    Vockin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvett, V.,
    Expelling, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
    A Functional Genomics Program for Soybean (NSF 9872565)
    Other ESTs: AW153139 corresponding to GM-1015-3315 (5')
    Contact: Vockin, L.O., PI, A Functional Genomics Program for
    Soybean (NSF 9872565)
    Lewin, H. A., Director, Keck Center for Comparative and Functional
    Genomics
    University of Illinois
    Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
    Tel: (217) 244-6147
    Fax: (217) 333-4582
    Email: l-vockin@uiuc.edu
    This clone is available through: Genome Systems, Inc. 4633 World
    Parkway Circle St. Louis, Missouri 63134. For further information
    call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
    427-3324 or contact: clones@genomesystems.com or info@genome
    systems.com web site: www.genomesystems.com
    Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
    Location/Qualifiers
    1..549
    /organism="Glycine max"
    /db_xref="taxon:3847"
    /clone="GM-r1070-2683"
    /clone_1lb="GM-r1070"
  
```

```

BASE COUNT      170 a      125 c      93 g      158 t      3 others
ORIGIN

Query Match
Best Local Similarity 39.6%; Score 163.8; DB 138; Length 549;
Matches 244; Conservative 0; Mismatches 112; Indels 3; Gaps 1;

OY 23 gggggaagaataatacagaatcgaagatgctacacgggtttatatacaccga 82
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 509 GAGAGACAAAGTTGTCAGAGAGTTCAAGATGTTGATCAAGTTTATATCAACAGA 450
OY 83 tttcttcgcgcgaaggtttgacacacacgttaacctgttatacacaacacagcc 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 449 TATTTCTTCTCGTGGCTTTTATATCAACAGTTAATTTGTCATCAATTAATCTTCC 390
OY 143 aatcaaatctgtacagaagatgaacctgtatctgaaggttacttgcacagaattggcag 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 389 AAATTAACACAGTTTACGGGATGACACAGATTATGAGTGTATGTCACAGGGTTGGCAG 330
OY 203 agctggcgcttggccggaaggtgctgtgtcaactgtgtgtgtgtaaacacgatta 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 329 AGCTGGCGCTTTGGCGCAAGGGCGCTGTATTAACTGATATGATGAATGATGA 270
OY 263 tgtgtgatgaagaagatcgaaagactatctccacagatgtgcctatgttgcgaattg 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 269 AAGGCTCATGTCGAAGATTGAACCAATTTGGCATTTGTGTAAGTGAAGTGCAGCT-- 210
OY 323 gcaacgtgaataataataatagagctgatacgaagatgcacgttactttaaagttacc 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 211 -CAAAGTGTGAAGATATTAAGCCGCTCTCAAGAGACGGGTTACTGCAATGATGCC 154

RESULT 9
BE131561 518 bp mRNA EST 21-JUN-2000
LOCUS BE131561 L48-1550r3 Ice plant lambda Uni-zap XR expression library, 48 hours
DEFINITION NCI treatment Mesembryanthemum crystallinum cDNA clone L48-1550,
    mRNA sequence.
ACCESSION BE131561
VERSION BE131561.1 GI:8578924
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
    Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
    Caryophyllales; Alstroceae; Mesembryanthemum.
    1 (bases 1 to 518)
  
```

/note="The library GM-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library GM-r1070. The cDNA clones of the rerecked GM-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the Laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html>. Rerecking was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3 sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/BioTech/Keck.html>. Note: The corresponding 5' EST from each clone in the GM-r1070 library is listed in the "OTHER EST" field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under "OTHER EST".

AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry and Molecular Biology
Oklahoma State University
350 Noble Research Center, Stillwater, OK 74078-3035, USA
Tel: 405-744-6207
Fax: 405-744-7739
Email: jcushman@biochem.okstate.edu

PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L48-16
Seq primer: T3
High quality sequence stop: 350
POLY-A-No.

FEATURES
source Location/Qualifiers

1. 518
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L48-1550"
/clone_lib="ice plant lambda Uni-Zap XR expression library"
/tissue_type="leaf, 48 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 142 a 83 c 155 g 138 t
ORIGIN

Query Match 37.4%; Score 154.8; DB 163; Length 518;
Best Local Similarity 66.5%; Pred. No. 2.3e-36;
Matches 222; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Oy 25 gggaaagataatacagaatcaagaatggtacacggggtctctatacaaccgag 84
Db 88 GAGACAGATAGTCAAGAGTTAAAGATGCTTGACTAGGCTTATATCAACTGATC 147
Oy 85 ttctgtcgcgaagttttgacacgaacacgcttaacctgtctcaactcagcagccaa 144
Db 148 TTTCTTCCAGGGGATTTGATCAATCAAGTCAATTTGGTGTCAACTATGACCTTCCAC 207
Oy 145 tcaaatctgttacaagaagatgaacctgattacgaagtgaacttcacagaatctggcagag 204
Db 208 TGAACATGATTAACGGGACAGCCTGACTGTGAGGTATACCTGATAGAGTAGGAGAG 267
Oy 205 ctgggcgctttggccggaaggtgctgtgtcaacttgcgtgtgtgtcaaacgagttatg 264
Db 268 CAGGCGCTTCCGGCGGTAAAGAGCGGCTGTTCATATTTGTTGTGGTATAGGACATCA 327
Oy 265 ttgtgatgaagaatcgaagaactttccacagacagtgccctatgttcggaattggc 324
Db 338 TGATATGAGAGATTCGGAAGCATTTTCGCTCGACATTCAGGAGGCTTCGTACG 387
Oy 325 aacgtgaaataatataatgaggtgatactccaagga 358
Db 388 ACAGGAGAGAGCATTTTGAGCGGCTCTCAAGAA 421

RESULT 10
A1486996 598 bp mRNA EST 29-JUN-1999
LOCUS A1486996
DEFINITION EST245318 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
CLD6H1, mRNA sequence.
ACCESSION A1486996
VERSION A1486996.1 GI:4382367
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

REFERENCE Magnoliophyta: eudicotyledons; core eudicots; Asteridae: euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 598)

AUTHORS Alcalá, J., Vrehlov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman
, C.L., Niernan, R., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley
, S.D. and Giovannoni, J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
CONTACT: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: df@rsche.clemson.edu

FEATURES
source Location/Qualifiers

1. 598
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLD6H1"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRP"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLD6 - Tomato Carpel EST library. OligodT-primed and
directionally cloned cDNA in vector Lambda Zap II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 196 a 101 c 148 g 153 t
ORIGIN

Query Match 37.4%; Score 154.8; DB 21; Length 598;
Best Local Similarity 65.1%; Pred. No. 2.4e-36;
Matches 228; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Oy 25 gggaaagataatacagaatcaagaatggtacacggggtctctatacaaccgag 84
Db 231 GAGACAAATATTCAGAGAGTTCAAGATGATGTGACACAAATTTTATATCACTGATC 290
Oy 85 ttctgtcgcgaagttttgacacgaacacgcttaacctgtctcaactcagcagccaa 144
Db 291 TTTCTTGTGCTGATGATTTGACCAATCAAGTTAATTTGGTGTATATATGACCTCCGTG 350
Oy 145 tcaaatctgttacaagaagatgaacctgattatgaggtgtacttcacagaatctggcagag 204
Db 351 TGAGACACAGAGTCCACACAGAGCCAGATCATGAGTATACCTGCATCGGATGGTAGAG 410
Oy 205 ctgggcgctttggccggaaggtgctgtgtgtcaacttgcgtgtgtgtgtcaaacgagttatg 264
Db 411 CAGGAGCTTTTGAGCGCAAGGCTGATATTCATCTGCTGTGACGTGACAGACAAACA 470
Oy 265 ttgtgatgaagaatcgaagaactttccacagacagtgctctatgttcggaattggc 324
Db 471 TGCTATATGCAAGATTTGAGACATTTTCACATCAAGTGGCTGAGATTCCTTATGGA 530
Oy 325 aacgtgaaataatataatgaggtgatactccaaggaatgacagttactctta 374
Db 531 GTATGAGAGAGATTTGCAAAATGCTGTGAGATAGTCTGGCTGTAA 580

RESULT 11
BG101919 484 bp mRNA EST 30-JAN-2001
LOCUS BG101919
DEFINITION RHIZ2_21_E02.g1_A003 Rhizome2 (RHIZ2) Sorghum prolingum cDNA, mRNA
sequence.
ACCESSION BG101919
VERSION BG101919.1 GI:12616654
KEYWORDS EST.
SOURCE Sorghum prolingum.

ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 484)
AUTHORS Cordonier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonier-Pratt MM
Department of Botany
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: PolYTmX
High quality sequence start: 6
High quality sequence stop: 480
POLY=yes.

FEATURES
Location/Qualifiers
source 1..484
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_1id="Rhizome2 (RHIZ2)"
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector Lambda Zap II.
Clones to be sequenced were prepared by mass excision."

BASE COUNT 108 a 102 c 122 g 152 t
ORIGIN

Query Match 37.0%; Score 153; DB 173; Length 484;
Best Local Similarity 91.5%; Pred. No. 8.1e-36;
Matches 173; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 209 ggcgttggccggaagatgctgtgtcaactgtgtgtgtgcaacgattatgttgc 268
|||||
Db 1 GCGCTTGGCCGGAAAGGCTGCTGTCTCACTTCCTGTGTCTCAACAGATTAATGTGT 60
|||||

QY 269 gatgaagaagatcgaagactattccagcacagtgctcctatgttcggaattggcaacg 328
|||||
Db 61 GATGAAGAAGATGTGAGGCTATTCCAGCACAGTACTAGGTGGGATTTGGCAACG 120
|||||

QY 329 tgaataataatagagctgattcctaagagatgcagcttactttaagttacacgaa 388
|||||
Db 121 TGAAGAAGATTTTGAAGCAGCTCTCAAGAGATGAGGTTTAC-TTTAAGTAACCAACCGAA 179
|||||

QY 389 ctgtgtgtg 397
|||||
Db 180 CTTGTGCTG 188
|||||

RESULT 12
AM933319 459 bp mRNA EST 30-MAY-2000
LOCUS EST359162 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone CLEF52L17 5', mRNA sequence.
ACCESSION AM933319
VERSION AM933319.1 GI:8108720
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 459)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ann,S., Rinning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)

COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU

FEATURES
Location/Qualifiers
source 1..459
/organism="Lycopersicon esculentum"
/cultivar="RA496"
/db_xref="taxon:4081"
/clone="CLEF52L17"
/clone_1id="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; CLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"

BASE COUNT 145 a 80 c 112 g 122 t
ORIGIN

Query Match 35.7%; Score 147.8; DB 121; Length 459;
Best Local Similarity 64.0%; Pred. No. 3e-34;
Matches 240; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 25 gggaaagataatcagaatcaagaatgctacaccgggttctatatacaccgatg 84
|||||
Db 60 GAGCAAAATATATCAACAGAGTTCAAAATGGATGACACAAATCTTATATCAACTGATC 119
|||||

QY 85 ttctgtctgaggtttgaccacagacagcttaacctgtcatcaactaactgaatgcac 144
|||||
Db 120 TTTCTTGCTCGGGGATTTGACCAATCAACAGTTAATTTGGTGTAAATTAATGACCTCCCTG 179
|||||

QY 145 tcaaatctgtgtacagaagatgacattgattatagtgatctgcacagaattggcagag 204
|||||
Db 180 TGACACATNTAGAGATCCAAACAGACGACATCATAGATATCTTCATCGGATTTGGTAGAG 239
|||||

QY 205 ctggcgcttggccggaagagtgctgttcaactctgtgtgtgtgtgtgtgtgtgt 264
|||||
Db 240 CAGGACGTTTTTGAAGCAAGAGTGCATATTCACCTTGTGTGACGACGACCAACA 299
|||||

QY 265 ttgtgagaagaagatcgaagactattccagcacagtgctgacctatgttcggaattggc 324
|||||
Db 300 TGTATATGTCAAAAGATTGAGAACCATTTTCAACATCAAGTGGCTGAGAT---CTCATGGA 356
|||||

QY 325 aacgtgaaaaatattatagagctgattcctaagagatgacglttactttaagttacacg 384
|||||
Db 357 ATAGTGAAGAAGACTTTGAATAATGCTCTGAAGATAGAGCTGGCTGCTGAAGCCGATATG 416
|||||

QY 385 gaaactgtgtgtgt 399
|||||
Db 417 AATGTGTGTATATT 431
|||||

RESULT 13
BE437127 578 bp mRNA EST 24-JUL-2000
LOCUS BE437127
DEFINITION EST408245 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLEG35H9, mRNA sequence.
ACCESSION BE437127
VERSION BE437127.1 GI:9434970
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids

	TITLE	I; Solanales; Solanaaceae; Solanum; Lycopersicon.
	JOURNAL	1 (bases 1 to 578)
	AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craen,M.B., Bowman,C.L., Romning,C.M., Niernm,W., Frazer,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley .S.D. Generation of ESTs from tomato fruit tissue, breaker stage Unpublished (2000) Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 4366 Fax: 864 656 4293 Email: df@frisch@clemson.edu 5 prime sequence.
	FEATURES	Location/Qualifiers
	source	1..578
		/organism="Lycopersicon esculentum"
		/cultivar="TA96"
		/db_xref="taxon:4081"
		/clone="CLEG35H9"
		/clone_1id="tomato breaker fruit, TIGR"
		/tissue_type="Pericarp"
		/dev_stage="breaker"
		/lab_host="SOLR"
		/note="Vector: pBluescriptSMKcMcdapl; Site_1: EcorI; Site_2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end o the fruit). Fruit were cut in half and the seeds and o locules were discarded prior to freezing the pericarp."
	BASE COUNT	185 a 96 c 141 g 156 t
	ORIGIN	
	Query Match	35.7%; Score 147.8; DB 167; Length 578;
	Best Local Similarity	64.0%; Pred No.3-ze-34;
	Matches 240; Conservative	0; Mismatches 132; Indels 3; Gaps 1
OY	25	gggaaagaatacaaggaatccaagtctacaccgcgggttcctaacaaccgatg 84
Dd	160	GAGACAAATATTCAAGAAGCTTCAAAGATGGATTGCACAATAATTCTTATACACTGATC 219
OY	85	tctctcgtagtgatttcgaaccaaacaccttaacctgtcatcaactacgacatgccaa 144
Dd	220	TTCTGTTCGGCGGAATTTGCCAATCACAGGTTAAATTTGGTGTTAATTATCATCCCTCG 279
OY	145	tcaaatttgtcacaggagatatgaacctgatatgaagtgactgtccagaatttggaaag 204
Dd	280	TGAGCAATGAGAGTCCCACAGACCAGATCAAGAGTTAATCTGCATTCGATTGGTAGAg 339
OY	205	clyggcgcclttgcgcgaaaagg tgcctgtctcaaatctgtctgtgtgtcaaacogattag 264
Dd	340	CAGGACGTTTTTGAGAGCAAGGTGCGATTTTCAACTGCTGTGACAGTACCCGAGAACAA 399
OY	265	ttgtgatgaagaagatcgaagaactatlccacgacagatgltgccttatgttcgnaattggc 324
Dd	400	TGCTAATGTCANAAATTTGGAACCACTTTTCAACCAATCAAMTGGCTGAGAT---CTCATGGA 456
OY	325	aacggaaaaaratatgatagsgctgatatcaagaagatgcaagtttactttaagttaeccag 384
Dd	457	ATVAGGAAGAGACTTTGAAAATGCTGTGAADAATGACTGCTTGTGTGTGAAGCCGAGTATG 516
OY	385	gaacctgtgtgtct 399
Dd	517	AATGTTGGTTAATT 531
RESULT 14		
LOCUS	AI782217/c	647 bp mRNA EST 29-JUN-1999
DEFINITION	EST263096 tomato susceptible, Cornell Lycopersicon esculentum CDNA	
		Clone CLESI8B13, mRNA sequence.

ACCESSION	A1782217
VERSION	A1782217.1
KEYWORDS	GI:5280258
SOURCE	EST.
ORGANISM	Lycopersicon esculentum Tomato.
TITLE	Lycopersicon esculentum Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; solanaceae; Solanum; Lycopersicon. 1; (bases 1 to 647)
REFERENCE	D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Updon,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman, C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tankley,S.D., Giovannoni,J.J. and Martin,G.B. Generation of ESTs from Pseudomonas susceptible tomato unpublished (1999) Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU 5 prime sequence.
JOURNAL	Location/Qualifiers
COMMENT	1..647 /organism="Lycopersicon esculentum" /cultivar="R11-13 (Rio Grande x Money Maker)" /db_xref="taxon:4081" /clone="cLESI8B13" /clone_lib="tomato susceptible, Cornell" /issue_type="leaf" /dev_stage="4-week old" /lab_host="SODR" /note="vector: pBluescript SK(-); Site_1: EcoRI, Site_2: XhoI; cLES - Tomato Pseudomonas Susceptible EST library, directionally cloned cDNAs inserted into pBluescript SK(-)) at 5' end with EcoRI and 3' end with XhoI site"
BASE COUNT	182 a 147 c 120 g 198 t
ORIGIN	
Query Match	34.9%; Score 144.6; DB 24; Length 647;
Best Local Similarity	63.5%; Pred.No.3.le-33;
Matches 238; Conservative	0; Mismatches 134; Indels 3; Gaps 1;
OY	25 gggaagaataacaaagatcaagaattggtacaccgggtcttataccaacgatg 84
DB	564 GAGCCAAATATATCAAGCATTTCAAAGATGGATTGCACAAATTTCTTATATCAACTGATC 505
OY	85 ttcttgctcgaggtttcgaaccaacacgcyttaacctgtgtcatcaactacgaacatgccaa 144
DB	504 TTCTTGCTCGCGGGAATTTGCACAACAGGTAAATTTGGTGGTTAATTAATGACCTCCCG 445
OY	145 tcaaatttggtacaagaagatatgaacctgattatgaaggttactgtcacgaagtgtgcagag 204
DB	444 TGAGACATGAGAGTCCAACACAGAGCACAGATTCACAGAGATAACTTGGATCGGATGGTAGG 385
OY	205 ctggggcgcttgcccggaagaagtgtctgttcaacttgctgtgtgtgttcaaacgattatg 264
DB	384 CAGCACGTTTTTGACGCCAAGGTGGGATATTAACTTGGCTGTGCAGTGCACGACGACMACA 325
OY	265 ttgtgtagaagaagatcgaagaagatttcacgacacagltgtgcttatgtltycgaatltygc 324
DB	324 TGCATATGTCANAAGATTGAGAACCATTTCACACATTCACAAAGTGCGTAGAT---CTCATGGA 268
OY	325 aaactgaaaaatatatcatgaagctgatctccaagaatgacgaagttactttaaagttccacag 384
DB	267 ATATGAGAGAAACACTTGTGAATAATGCTGGAACATAGCTGGCTGGTATGACCGAGATATG 208
OY	385 gaactgtgtgtgtc 399
DB	207 AATGTTGGTTATATTT 193

RESULT 15
AI777083

LOCUS
DEFINITION

AI777083 593 bp mRNA EST 29-JUN-1999
EST252050 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEL119 similar to similar to eukaryotic initiation factor, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AI777083
AI777083.1 GI:5275065
EST.

ORGANISM

tomato.

Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons: core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE

1 (bases 1 to 593)

AUTHORS
TITLE
JOURNAL
COMMENT

Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Roming,
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfisch@clemson.edu

FEATURES

source

Location/Qualifiers
1..593

/organism="Lycopersicon esculentum"
/cultivar="TA96"
/db_xref="taxon:4081"
/clone="cLEL119"
/clone_id="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST library"

BASE COUNT 195 a 96 c 137 g 163 t

ORIGIN

Query Match 33.5%; Score 138.8; DB 24; Length 593;
Best Local Similarity 67.8%; Pred. No. 1.8e-31;
Matches 194; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```

QY 25  gggaaagataatcagaatccaagaatgctacaccggggttcctatcatcaaccgatg 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90  GAGACAAATATATCAAGAGATTCAAGATGATGACACAATCTTATATCAACTGATC 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 85  ttcttgctcgaggtttgaccagaacacglttaacctgtcatcaactacgacatgcca 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150  TTTCTTGCTCGTGATTTGACCAATCACAGTGAATTTGGTGAATTAATGACCTCCCTG 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145  tcaaatgttgtaagaagatgaactgtataggtgttacttgacacagaattggcagag 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210  TGAGACACGAGAGTCCACACAGACGACGATCATGAGTATACCTGCATCGATTGGTAGAG 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 205  ctgggagcttggcggaaggtgctgttcaactgtgtgtgtcaaccgatattatg 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270  CAGGACGTTTGGACGCAAGGTTGCGATATTCACACTTCTGTGCAAGTACAGACACA 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265  ttgtgatgaagaagatcgaaactatttccagacagtggtgacctta 310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330  TGCTAATGTCAAGATGAGAACCATTTCAACCATTCACAGTGCCTGA 375

```

Search completed: May 8, 2001, 05:34:35
Job time: 9044 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:57:34 ; Search time 2791.02 seconds
(without alignments)
1759.385 Million cell updates/sec

Title: US-09-619-643-5

Perfect score: 333

Sequence: 1 tgataactctgtccatcctt.....acattacatgagaaagtgc 333

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_higo_hum:*
- 20: em_higo_inv:*
- 21: em_higo_rod:*
- 22: em_hig_hum1:*
- 23: em_hig_hum2:*
- 24: em_hig_hum3:*
- 25: em_hig_hum4:*
- 26: em_hig_hum5:*
- 27: em_hig_hum6:*
- 28: em_hig_hum7:*
- 29: em_hig_hum8:*
- 30: em_hig_inv1:*
- 31: em_hig_inv2:*
- 32: em_hig_other:*
- 33: em_hig_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_v1:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_v11:*
- 59: gb_v12:*
- 60: gb_hig1:*
- 61: gb_hig2:*
- 62: gb_hig3:*
- 63: gb_hig4:*
- 64: gb_hig5:*
- 65: gb_hig6:*
- 66: gb_hig7:*
- 67: gb_hig8:*
- 68: gb_hig9:*
- 69: gb_hig10:*
- 70: gb_hig11:*
- 71: gb_hig12:*
- 72: gb_hig13:*
- 73: gb_hig14:*
- 74: gb_hig15:*
- 75: gb_hig16:*
- 76: gb_hig17:*
- 77: gb_hig18:*
- 78: gb_hig19:*
- 79: gb_hig20:*
- 80: gb_hig21:*
- 81: gb_hig22:*
- 82: gb_hig23:*
- 83: gb_hig24:*
- 84: gb_hig25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_r01:*
- 95: gb_r02:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	71.6	21.5	110848	12 AC007017	AC007017 Arabidops
2	69.6	20.9	115175	12 AC002505	AC002505 Arabidops
3	41.6	12.5	174958	72 AC055758	AC055758 Homo sapi
4	40	12.0	142203	76 AC079521	AC079521 Mus muscu
5	40	12.0	181655	71 AC026946	AC026946 Homo sapi
6	40	12.0	207411	77 AC084746	AC084746 Mus muscu
7	39.6	11.9	206996	83 AP002985	AP002985 Homo sapi
8	38.6	11.6	1440	6 DG87812	DG87812 Drosophila
9	38.6	11.6	151560	72 AC036223	AC036223 Homo sapi
10	38.6	11.6	174217	75 AC073167	AC073167 Homo sapi
11	38.4	11.5	33714	66 AC020137	AC020137 Drosophila

C 12	38.4	11.5	262205	4	AE003647	AE003647 Drosophila
C 13	38.4	11.5	316756	6	DROSADH06	AE003612 Drosophila
C 14	38.2	11.5	11681	3	U32813	U32813 Haemophilus
C 15	38.2	11.5	105576	91	HS106120B	281369 Human DNA s
C 16	38.2	11.5	110000	80	AL357872-0	AL357872 Homo sapi
C 17	38.2	11.5	138859	90	AL359076	AL359076 Human DNA
C 18	38.2	11.5	149321	91	HS106120	AL072295 Human DNA
C 19	38.2	11.4	136949	78	AL138771	AL138771 Homo sapi
C 20	38	11.4	175543	75	AC073378	AC073378 Homo sapi
C 21	37.8	11.4	162760	86	AC007969	AC007969 Homo sapi
C 22	37.4	11.2	3090	94	AF164632	AF164632 Mus muscu
C 23	37.2	11.2	5200	3	SYNOLIVE	U29131 Synecocyst
C 24	37.2	11.2	122056	2	D90902	D90902 Synecocyst
C 25	37.2	11.2	131227	78	AL136098	AL136098 Homo sapi
C 26	37.2	11.2	152228	89	AL135902	AL135902 Human DNA
C 27	37	11.1	1443	9	AX004415	AX004415 Sequence
C 28	37	11.1	2095	9	AX004413	AX004413 Sequence
C 29	37	11.1	2119	91	H0517571	Y17571 Homo sapien
C 30	37	11.1	2368	89	AK025779	AK025779 Homo sapi
C 31	37	11.1	2913	88	AK000766	AK000766 Homo sapi
C 32	37	11.1	3150	88	AF118838	AF118838 Homo sapi
C 33	36.8	11.1	170219	77	AC079910	AC079910 Homo sapi
C 34	36.8	11.1	187295	67	AC021531	AC021531 Homo sapi
C 35	36.8	11.1	197371	80	AL355524	AL355524 Homo sapi
C 36	36.6	11.0	98053	12	AC005292	AC005292 Genomic s
C 37	36.6	11.0	123462	78	AF207954	AF207954 Homo sapi
C 38	36.6	11.0	131761	92	HS3493D19	AL096868 Human DNA
C 39	36.6	11.0	163594	67	AC021924	AC021924 Homo sapi
C 40	36.6	11.0	171053	70	AC025675	AC025675 Homo sapi
C 41	36.6	11.0	180947	81	AL445530	AL445530 Homo sapi
C 42	36.6	11.0	185855	87	AC022542	AC022542 Homo sapi
C 43	36.4	10.9	930	94	AB005143	AB005143 Rattus no
C 44	36.4	10.9	930	94	AB006613	AB006613 Rattus no
C 45	36.4	10.9	12222	93	H0MA1ATP	K02212 Human alpha

ALIGNMENTS

RESULT 1	AC007017	110848 bp	DNA	PLN	05-APR-2000
LOCUS	AC007017	Arabidopsis thaliana chromosome II section 196 of 255 of the complete sequence. Sequence from clones T20F21, F11F19.			
DEFINITION	AC007017	AE002093			
ACCESSION	AC007017.3	GI:6598647			
VERSION					
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.				
AUTHORS	1 (bases 1 to 110848) Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldblyum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffitt, K.S., Cronin, L.A., Shen, M., Vanden, S.E., Umayam, L., Talon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Cressy, T.H., Goodman, H.M., Somerville, C.R., Coppenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.				
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana				
JOURNAL	Nature 402 (6763), 761-768 (1999)				
MEDLINE	20083487				
PUBMED	10617197				
REFERENCE	2 (bases 1 to 110848)				
AUTHORS	Lin, X.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA				
COMMENT	On Dec 17, 1999 this sequence version replaced gi:4510360. The sequence and annotation of chromosome 2 were merged from those				

of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://artur.epm.ornl.gov/pub/xgrail>), GeneFINDER (Phil Green, University of Washington), GenScan (Chris Burge, <http://genome.stanford.edu/GENSCAN/HTML>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curation based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/Mashu/ABI consortium for sequencing BAC clones F6F23, F536, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	Location/Qualifiers
source	1..110848 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II" complement(1..2225) /note="Sequence from clone T20F21"
misc_feature	complement(join(303..579,728..1200,1278..1363,1895..1987,2074..2169,2258..>2588)) /gene="At2g35770"
gene	complement(303..>2588) /gene="At2g35770" complement(join(315..579,728..1200,1278..1363,1895..1987,2074..2169,2258..2588)) /gene="At2g35770"
CDS	/codon_start=1 /product="putative serine carboxypeptidase II" /protein_id="AAF18665.1" /db_xref="GI:6598648"
misc_feature	/translation="MVIALDLYVSSDDAKEDKMKDKITSLPGOPNLNFSOGSYVLPDAGRALFWLTETAPRSPGKPLVLTSSDELTVGCKRFGEDAYRFLVRLMKEPEYKERAFLVAGESYAGHYIPELAQLIVNNKKGKNTINLKGILMGNPLVDYDNDKGMRDYWNHGLIDSEYNDLTKCLNDLSIEPKINCAALNQAISEFGDIDPYNINSPACTTHASNNEMQAMVYRGNDCEVGYTRYKMDPNVHKSFHARLNGSTPWTCSRYIRKNWKDSPKMLPITKIKLQAHLRITWFSGSDSAVLPLSTGRHSINMKLKSSKRMYPWHSHGLVGSMSQYVERGLTYTTRVAAAGHEVPLSPRLALFLFTHLANHSLPSSS"
misc_feature	complement(2226..>110848) /note="Sequence from clone F11F19"
misc_feature	complement(join(3802..4009,4095..4200,4288..4383,4469..4596,5049..5315,5392..5465,5538..5630,5706..5801,5883..>6240)) /gene="At2g35780"

```

gene      complement(3802..>6240)
           /gene="At2g35780"
           /note="F11F19.31"
           complement(join(3869..4009,4095..4200,4288..4383,
           4469..4596,5049..5315,5392..5465,5538..5630,5706..5801,
           5883..6240))
           /gene="At2g35780"
           /codon_start=1
           /product="putative serine carboxypeptidase II"
           /protein_id="A02147.1"
           /db_xref="GI:4510391"
           /translation="MARLLLFEEFLILLHYASCSSRHEQKDRIFHLPGEPNDYFESH
           EFGYIVNESAGRALFYWLTPESPSPENESKPLVNLNGGCGSSVYGAEEIGPER
           INPDGKLYHNPSYMNKLANLFLSPAGVGFSTNPSDLYTAQRDQRTADYVFLY
           KNEFRPQYKHREPIYAGESTYAGHYVQLSQIYEKRNPAINFGFJVGNAVDDIHD
           YVLEFEYMAHGLISDLYHNLRITCEGSEHPSKCTAMEAADLEQGNIDYSLY
           TWCKKEAALRSRSRVRHMMRAADPCTEKSGMVFNSPEYOKAMHNIIGLAP
           MKGCSDIVGEKWDSPSLMLPIYKELIAGLRITWFGDTSVVPITGTRSTRALK
           OPLSKWYPMNDQGVGMSQVYKGLTLVTIHGAGHEVRLFRPRARFLPLDNRKL
           PM"
CDS
           6179..6231
           /rpt_family="GAA)n"
           6634..6659
           /rpt_family="POLY_A"
           join(<6871..6943,7469..8257)
           /gene="At2g35790"
           <6871..8257
           /gene="At2g35790"
           /note="F11F19.30"
           join(6871..6943,7469..8112)
           /gene="At2g35790"
           /note="unknown protein"
           /codon_start=1
           /protein_id="A021478.1"
           /db_xref="GI:4510390"
           /translation="MGRSALIHLSRSRRLSSFTGYGHHRSIAGSSSVNPKVR
           FOYPSLSORSMASGATREDDEHKIISIPQEKKEKDGQVYGFISSTIKVKLL
           SLSCCSVSLGPIYTPMSPGLNIMKGAVASVITLSSTRALHWPSPYHKKR
           WQPSDTFEYEMTWLCTFSAKTLKESDIRIPDQRYVSKADGNFVADADCPK
           ALLARLTPPKDAHDSAFKNL"
repeat_region
           complement(10032..10115)
           /rpt_family="(TAA)n"
           10128..10186
           /rpt_family="POLY_A"
           complement(10214..10308)
           /rpt_family="(GA)n"
           join(<10573..11707,12096..12513,12738..12882,13090..13205,
           13295..13436,13827..14060,14144..>14488)
           /gene="At2g35800"
           <10573..>14488
           /gene="At2g35800"
           /note="F11F19.29"
           join(10573..11707,12096..12513,12738..12882,13090..13205,
           13295..13436,13827..14060,14144..14488)
           /gene="At2g35800"
           /codon_start=1
           /product="putative mitochondrial carrier protein"
           /protein_id="A021477.1"
           /db_xref="GI:4510389"
           /translation="MYSKNDHIEFLFNSIOLVKDVLPIELGVKKARQDIENCMWIK
           ERDLGLYRSGSRNRKRRIYATPEFDNATNNOCVYVDERRKGLSIRKIPYSLG
           MSPNLYSEKLRGNDVYVAKKDSLEKKDDSDTDCPKAMMISLPGSVHAPRP
           FKICKRIHKMGDDENSLRKHLGSKSAAPYSRKRVRCOSVSEYKESNPSTIEAVCF
           VEMLAONLQKLDQFIODSSSESCSKSEASSNSDPLFNWKRKLDVNGFLNLF
           ARVDVAGSIGGLTSHISEGDESSTAGSESAVSPONLATGLSLPSLSENRK
           STLSITLTELIELPOLGRPSRDPKRLISVDFFRYTESGRFFELDDGCG
           KVLLEDEIAMRRRLPRYAKEFMRARSHLSVDFSGMOFLSIMOKETILRAYT
           SLCTKSGTLKSSILASLNNAGLPANEENAIAMRRKADTESISYGHRRNMVLL
           PYRLQDDPRNIREAFYVAVAPVALPADVYKSLAGGLASALSTSLMHPIDITK
           TRVASTLSFEPIVIAKLPEIGRVGVYGSIPAILGQPSHGLRGTIEASKVLINFA
           PNLEIIVITLVSFGMRDSNFVQVOSIASCSTLGTAVRICEVYKVLQVLOG
           MFNNVGAIVGTWKDQSGFRTGTATLREVLVYVGMGLVSESKMAQOAGRL
           EAMETIAGVAVSGIAIVTTPDVMTKMTATPGRPISSMNVVVSILRNHGPLGLF

```

```

mRNA      KGAVPREFVWAPLGAAMFAGYELAKKAKMKNEDAVLADLGOKKLC"
           join(15316..15481,15707..>16058)
           /gene="At2g35810"
gene      15316..>16058
           /gene="At2g35810"
           /note="F11F19.28"
           /note="F11F19.28"
           join(15330..15481,15707..16058)
           /gene="At2g35810"
           /note="unknown protein"
           /codon_start=1
           /protein_id="A021476.1"
           /db_xref="GI:4510388"
           /translation="MAKSPVEVNIIPLEATPEPFAEYGVTEASRDAGYGPNDALD
           LSKGIPRLYLRLKEPLGFTKTHAKKQVCLSGDITWGVKRPSTLEDDGR
           RVDIVKAKSCHLTIPEVEIRFRTSGPRFVVLKLRHTWAGLFLSGSSIMDPTNLEL
           SNNVNSQ"
           join(17058..17220,17618..>18005)
           /gene="At2g35820"
           17058..>18005
           /gene="At2g35820"
           /note="F11F19.27"
           join(17069..17220,17618..18005)
CDS
           21.5%: Score 71.6; DB 12; Length 110848;
           Query Match Best Local Similarity 69.0%: Pred. No. 2.3e-09;
           Matches 98: Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy      1  tgataatcttgccatccttggaacgaggggccacttgagctctttaaggcgcgatcc 60
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      14301 TGGTTGTTGTAATCATCTCTGCGTAATAGAGGACCGCTAGTGTGTTGTAAGACAGACATCC 14360
Qy      61  ccggctcttcttgatctcctcttggaacgaacttgcagcgctatagcttgcca 120
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      14361 CGAGGTTCTTGCGTGCTCTCTAGCTGTAAGACTTGTGCGTACGACACTAGCA 14420
Qy      121  agaagcaatgataaagatga 142
           ||| ||| ||| ||| |||
Db      14421 AGAAGCATATGCAGAGAACGA 14442
RESULT  2
AC002505/c
LOCUS   Arabidopsis thaliana chromosome II section 148 of 255 of the
DEFINITION
ACCESSION AC002505 AEO02093
VERSION    AC002505.2 GI:6598374
KEYWORDS   HNC.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
           Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
           Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
           Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 115175)
AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
           Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
           Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
           Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., Vanaken,S.E., Umayam,L.,
           Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
           Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
           Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
           Venter,J.C.
           Sequence and analysis of chromosome 2 of the plant Arabidopsis
           thaliana
           Nature 402 (6763), 761-768 (1999)
JOURNAL  MEDLINE
PUBMED   20083487
10617197
REFERENCE 2 (bases 1 to 115175)
AUTHORS   Lin,X.
TITLE     Direct Submission
COMMENT   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA
           On Dec 17, 1999 this sequence version replaced gi:2739359.

```

The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tdb/ac/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<ftp://archur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NeuplanGene (<http://www.cds.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arjan Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F536, T17A5, and T13L16, the ESSA group for sequencing clone F13J4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Saeoshil Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Liklin Zhou, Hanif Khailak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

Location/Qualifiers

FEATURES
source

```
1. .115175
  /organism="Arabidopsis thaliana"
  /cultivar="Columbia"
  /db_xref="taxon:3702"
  /chromosome="II"
  complement(1. .1362)
misc_feature
  /note="5' end of GB:D83257 Erecta locus"
  1. .>115175
  /note="Sequence from clone T9J22"
  join(<919. .1167,1274. .1322,1424. .1494,1584. .1678,
1757. .1824,1896. .>2145)
  /gene="At2g26340"
  <919. .>2145
  /gene="At2g26340"
  /note="T9J22.1"
  join(919. .1147,1274. .1322,1424. .1494,1584. .1678,
1757. .1824,1896. .2145)
  /gene="At2g26340"
  /note="unknown protein"
  /codon_start=1
  /protein_id="AAC14486.1"
  /db_xref="GI:2739360"
  /translation="MLNVSHNNLSNGDKTPVLSPMSSLHATVSLPFSNSHKKL
TCAVLAIPPWKQSRRIIVSFFLSRLLLPNDAMAGLMOKLKYVRKKADPLEAY
VPAVLAQIDLEGFLNVEKFEFACRQRLSPASSLRVIRAVAYADDFSK
TATDVDCRLALEELDSLFELASRKSANATVLMKSGCTALFALDSLQVPSQVL
DKGRAMVEYRSASEEDASDLSSESEIKOSII"
complement(join(<3432. .3534,3781. .3857,4277. .4371,
4623. .4756,4846. .4924,4995. .5064,5159. .5348,5428. .5522,
5605. .5801,5900. .6081))
  /gene="At2g26350"
  complement(<3432. .6081)
  /gene="At2g26350"
```

gene

CDS

```
/note="T9J22.2"
complement(join(3432. .3534,3781. .3857,4277. .4371,
4623. .4756,4846. .4924,4995. .5064,5159. .5348,5428. .5522,
5605. .5694))
  /gene="At2g26350"
  /codon_start=1
  /product="putative peroxisomal membrane protein"
```

```
/protein_id="AAC14514.1"
/db_xref="GI:3075382"
  /translation="MKLLGOMLYVLTITGSGQQTGSEYCDIIQVAGPYGSPTPAR
ALFIYQAVPIYAIARISTRATQAVTFDESDDEFQSDHISPRMIDLPSSOYEST
SVSRNLNRLMBSMRHAIORHPVLPVAREVLOLVRLNMLFPEGFYHISRASG
VRVYFIGQLNRPVYQILGVFLIQLCILAEGIRRSLSITSSIOQASIGYOTS
GGRLPVINEBSNLITSEAEKGNMSTSTSTSEAVGKCTCLSTRQHPATPCGHVFC
MSCIMWENKQECPLCRTPTHTSSVLCVLSDF"
```

mRNA

```
complement(join(6436. .7591,7666. .7899,8348. .8453,
8524. .>8771))
  /gene="At2g26360"
```

gene

```
complement(6436. .>8771)
  /gene="At2g26360"
```

CDS

```
/note="T9J22.3"
complement(join(7268. .7591,7666. .7899,8348. .8453,
8524. .8771))
  /gene="At2g26360"
```

```
/codon_start=1
  /product="putative mitochondrial carrier protein"
```

```
/protein_id="AAC14487.1"
/db_xref="GI:2739362"
  /translation="MVNKAARKRHVSPESLILSLINDLLPINSOKLDCSPRRKSNSDA
CSFIYSLNPAKVNFTSRNSNEIAPLINAIVAKKNGKDFGTGVDGHLKLSALAGI
SCAFSAFLMHPDVTYKVSINSISFTYVIGTTLRTICEVLRQLRQANQDNTVEATYST
WHQESLKGIFRGVTLREVPFYVAGGLYNQSKVYERQDGRLEPWEIIVAGALS
GGFVAVITTPEDVITKRMVTAPQGVLSMLAAVYSILTHEGPLAFYKGVAFRFTWAP
LGALNLAGEVELQKAMITPILNOSVHSE"
```

repeat_region
repeat_region

```
9205. .9321
  /rpt_family="POLY_A"
  complement(9750. .9784)
  /rpt_family="(TAAGA)n"
```

mRNA

```
join(<10301. .10400,11069. .11188,11275. .11450,
11592. .>11717)
  /gene="At2g26370"
```

gene

```
<10301. .>11717
  /gene="At2g26370"
```

CDS

```
/note="T9J22.4; predicted by genscan"
join(10301. .10400,11069. .11188,11275. .11450,11592. .11717)
  /gene="At2g26370"
```

```
/note="hypothetical protein"
  /codon_start=1
```

```
/protein_id="AAC14488.1"
/db_xref="GI:2739363"
```

```
/translation="MAISHAQPVLVLLLSLYLFAFGAGAVNFENCKNAPIDYNGIT
NVTRVEISPYVGPYDEPTITISGFTSDSDSYIIYATLHVLYKYENWYTIINDLSD
VNGEDPDSIEPEGEKFLVLRKVPGLQSLPHKDKSKIVISLVDEGDGDAEVLKMKVE
FDNPAPTTUSA"
```

mRNA

```
complement(<12280. .>13722)
  /gene="At2g26380"
```

gene

```
complement(<12280. .>13722)
  /gene="At2g26380"
```

CDS

```
/note="T9J22.5"
complement(12280. .13722)
  /gene="At2g26380"
```

```
/codon_start=1
  /product="putative disease resistance protein"
```

```
/protein_id="AAC14512.1"
/db_xref="GI:2739389"
```

```
/translation="MNSCSFINFEITFAVIFLRCLNPTAATCHPDDPAGILSPKSGI
TKDPSGIIITKRGKGTDCSSNMVSCPNKNNRVVLTIRIESDPAGIFLSGTISPLAKL
OHLSEGVATINKNITGPPPLFLPHLKLKYYILEATRLUSGLPANIGLNLIDLTVYK
IRFTGSIIPSSISNLTRLNLYNLKGLTGYILEAIAMLKLSINLIDGNRLSGTIPD
IEKSWNLITLISNRNRSGLKPLPSIAIAVLVLAFLGAINLNGSISPSYSRVALD
TLDSKNRSGAVPKSLAKIRKIANINSHLNLNPPVILNKKVYITITLIDSYKPFM
ETIPKMYASITGSLAKKGIKSLDDPKTRQDILVYSIDLSDNELSGSPILNPLKQ
AEQLKFRMSGNKLEFIDRLKLSFTTLETLDLSRNLVFGKVPARVAGLKTILNLSNHL
```

```

mRNA      CGKLPYTKFPESVYFAGNDCLGSPSLSHCKG"
           complement(join(<14752..15340,15776..>16156))
gene       /gene="At2g26390"
           complement(<14752..>16156)
           /gene="At2g26390"
           /note="t9j22.6"
           complement(join(14752..15540,15776..16156))
           /gene="At2g26390"
           /codon_start=1
           /product="putative serpin"
           /protein_id="AAC14489.1"
           /db_xref="GI:2739364"
           /translation="MELQSIENQNVNVARLAKVLETOVANGSNVYSPMSTNVLIS
           LTAGSNPVTKEITLSPMSPTDHLNAVLAKTADGCTERSDCLSTAHGWIDKSSY
           LKPSFELLLENSTKASCSOVDFAFKPEVEYIDENIVADVTNLIKQILSDCTDTIK
           EIRNSYLLILANAVYFKAASRKFDATKTDNPHLDGNVVKVPMKMSQDYLRGYD
           GFQVLRPLPYVEDRHFMSMYLYLPNDKGLAALLEKISTEPGFDSIHPIHRFPVDLR
           IPLNPSFEFKASEVLKDMGLTSPKSNLTVMVDSPSNGDKLHVSITIKKACIEVD
           EECTEAASVAIATMPOCLMRNDFVADHFLFTVRADNSGVLTFLGQVLDPSKH"
           repeat_region
           16923..16948
           /rpt_family="GAAAAn"
           repeat_region
           complement(17266..17299)
           /rpt_family="POLYA"

Query Match      20.93; Score 69.6; DB 12; Length 115175;
Best Local Similarity 68.68; Pred. No. 8.5e-09;
Matches 96; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 12 tccatccttgaacagaggtccacttgggtctttaaaggcgagatcccggtcttc 71
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7423 TCAATTCTCACACATGAGGCGCTCTTGCTTCTACAAAGACGCTGTTCCAGGTTTTC 7364

QY 72 tgaattgcctcttggagcaatgaacttgcaggtatagcttgcgaagagcaatg 131
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7263 TGGACTGCTCTCGAGGCTTGACTGCTATGACTGCTATGACTGCTATGACTGCTATG 7304

QY 132 ataaagatgagaaattc 151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 7303 ATCACACCTCTGAAATCATTC 7284

RESULT 3
AC055758 174958 bp DNA HTG 06-NOV-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-88H10, WORKING DRAFT SEQUENCE,
21 unordered pieces.
AC055758
AC055758.13 GI:11094550
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 174958)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Aisbrook,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbataia,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C.,
Burck,P., Burrell,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garcia,N., Gali,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollaway,C.,
Hollins,B., Homsj,F., Howard,S., Huber,J., Huiyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loussegod,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemko,S.,
Ogun,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plimms,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenkan,I., Rolfe,M.,
Riz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Stinson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Teitrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Ver,Y., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Wattlington,S., Williams,G., Williamson,A., Wiczcyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 174958)
Submitted (18-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 4, 2000 this sequence version replaced gi:9438586.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc_help@bcm.tmc.edu
----- Project Information
Center project name: HAZZ
Center clone name: RP11-88H10
----- Summary Statistics
Sequencing vector: MJ3, L08821
Chemistry: Dye-Primer Bodipy: 4% of reads
Chemistry: Dye-terminator Big Dye: 96% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 152739 bases at least Q40
Consensus quality: 164832 bases at least Q30
Consensus quality: 168697 bases at least Q20
Estimated insert size: 167965; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 4x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 25671: contig of 25671 bp in length
25672 25772: gap of unknown length
48259 48359: gap of 22488 bp in length
48360 48360: gap of unknown length
64585 64585: contig of 16226 bp in length
64586 64586: gap of unknown length
82038 82038: contig of 17355 bp in length
82138 82138: gap of unknown length
82139 82139: gap of unknown length
95060 95060: contig of 12922 bp in length
95061 95160: gap of unknown length
106868 106868: contig of 11708 bp in length
95161 106869: gap of unknown length
106869 118698: contig of 11730 bp in length
118699 118798: gap of unknown length

```

```
* 118799 126836: contig of 8038 bp in length
* 126837 126936: gap of unknown length
* 126937 136279: contig of 9343 bp in length
* 136280 136379: gap of unknown length
* 136380 144173: contig of 7794 bp in length
* 144174 144273: gap of unknown length
* 144274 150282: contig of 6009 bp in length
* 150283 150382: gap of unknown length
* 150383 155743: contig of 5361 bp in length
* 155744 155843: gap of unknown length
* 155844 158791: contig of 2948 bp in length
* 158792 158891: gap of unknown length
* 158892 162191: contig of 3300 bp in length
* 162192 162291: gap of unknown length
* 162292 164366: contig of 2075 bp in length
* 164367 164466: gap of unknown length
* 164467 167430: contig of 2964 bp in length
* 167431 167530: gap of unknown length
* 167531 169991: contig of 2461 bp in length
* 169992 170091: gap of unknown length
* 170092 171096: contig of 1005 bp in length
* 171097 171196: gap of unknown length
* 171197 172375: contig of 1179 bp in length
* 172376 172475: gap of unknown length
* 172476 173793: contig of 1318 bp in length
* 173794 173893: gap of unknown length
* 173894 174958: contig of 1065 bp in length.
Location/Qualifiers
1. 142203
/organism="Mus musculus"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-88H10"
BASE COUNT 55598 a 30494 c 30963 g 55825 t 2078 others
ORIGIN
Query Match 12.5%; Score 41.6; DB 72; Length 174958;
Best Local Similarity 53.0%; Pred. No. 0.55;
Matches 89; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 115 ttgcacgaagcagcaatgataaagaagagaaatccagtgatcaacacgagagatc 174
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 157849 ttgccatgacattgataatgattttaaataattttcaaatgcacagattttccca 157908
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 tgcacgaagcagtgatcaacatgcaacttttatttatttttgcggtcgc 234
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 157909 atgcgatttaattatgacattgacgtattagcttttcagtagcccatattaac 157968
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 agagagaacagacatctgcttttttgatatgggaatacatt 282
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 157969 taaatgcacaaatgacatcnaatatttttattttaaattattattat 158016
||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 4
AC079521 142203 bp DNA HTG 02-SEP-2000
DEFINITION Mus musculus clone Rp23-313W20, WORKING DRAFT SEQUENCE, 22
unordered pieces.
ACCESSION AC079521
VERSION AC079521.1 GI:9964886
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 142203)
AUTHORS Doe Joint Genome Institute.
JOURNAL Unpublished
TITLE Sequencing of Mouse
AUTHORS Doe Joint Genome Institute.
REFERENCE 2 (bases 1 to 142203)
AUTHORS Doe Joint Genome Institute.
TITLE Direct Submission
```

```
JOURNAL
COMMENT
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1857372
Center clone name: RPCI-23_313W20
-----
Summary Statistics
Consensus quality: 125899 bases at least Q40
Consensus quality: 134502 bases at least Q30
Consensus quality: 136312 bases at least Q20
Estimated insert size: 219000; agarose-fp estimation
Estimated insert size: 140103; sum-of-contigs estimation
Quality coverage: 5.52 in Q20 bases; agarose-fp estimation
Quality coverage: 8.63 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1410: contig of 1410 bp in length
1 1411 1510: gap of unknown length
* 1511 3323: contig of 1813 bp in length
* 3324 3423: gap of unknown length
* 3424 4558: contig of 1135 bp in length
* 4559 4658: gap of unknown length
* 4659 5830: contig of 1172 bp in length
* 5831 5930: gap of unknown length
* 5931 7613: contig of 1683 bp in length
* 7614 7713: gap of unknown length
* 7714 10210: contig of 2497 bp in length
* 10211 10310: gap of unknown length
* 10311 11472: contig of 1162 bp in length
* 11473 11572: gap of unknown length
* 11573 13839: contig of 2267 bp in length
* 13840 13939: gap of unknown length
* 13940 15864: contig of 1925 bp in length
* 15865 15964: gap of unknown length
* 15965 18962: contig of 2998 bp in length
* 18963 19062: gap of unknown length
* 19063 22563: contig of 3501 bp in length
* 22564 22663: gap of unknown length
* 22664 25886: contig of 3223 bp in length
* 25887 25986: gap of unknown length
* 25987 34910: contig of 8924 bp in length
* 34911 35010: gap of unknown length
* 35011 43064: contig of 8054 bp in length
* 43065 43164: gap of unknown length
* 43165 44643: contig of 1479 bp in length
* 44644 44743: gap of unknown length
* 44744 58525: contig of 13782 bp in length
* 58526 58625: gap of unknown length
* 58626 69408: contig of 10783 bp in length
* 69409 69508: gap of unknown length
* 69509 80063: contig of 10555 bp in length
* 80064 80163: gap of unknown length
* 80164 90270: contig of 10107 bp in length
* 90271 90370: gap of unknown length
* 90371 98772: contig of 8402 bp in length
* 98773 98872: gap of unknown length
* 98874 114715: contig of 15843 bp in length
* 114716 114815: gap of unknown length
* 114816 142203: contig of 27388 bp in length.
Location/Qualifiers
1. 142203
/organism="Mus musculus"
```

```

/db.xref=taxon:10090"
/clone="RP23-313M20"
/clone_id="RPCI mouse BAC library 23"
BASE COUNT 34179 a 34377 c 35406 g 36073 t 2168 others
ORIGIN

```

Query Match	12.0%	Score 40:	DB 76;	length 142203;
Best Local Similarity	57.0%	Pred. No.	1.5;	
Matches 73;	Conservative	0;	Mismatches 55;	Indels 0;
				Gaps 0;

Oy 88 gagcaatgaactcttgacgagctatgagcttgcagaaggcattataaaagtggaaaa 147
| | | | | | | | | | | | | | | | | |
Db 30400 GGCAGATGCAAGTCCTCAAAAAAATAAAATATAAATATAAATATAAATTA 30455

Oy 148 attccagtgagtcaacacgagagaactcttgcagaagaagcagtgatcaaccatgcaactt 207
| | | | | | | | | | | | | | | |
Db 30460 AATTAATAAACAACCAAAACCAACTTCATCTCCAAAGTTGTAGTGTTCATCAGTGCAACTAT 30519

QY	208	tttattt	215
Db	30520	TGAAGTGT	30527

RESULT	5
AC026946/c	
LOCUS	
DEFINITION	
AC026946	181655 bp DNA HTG 26-MAY-2000
	Homo sapiens chromosome 3 clone RP11-297C5 map 3, WORKING DRAFT
	SEQUENCE, 8 unordered pieces.

VERSION	AC026946.3	GI:8077058
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT	
SOURCE	human.	

REFERENCE
1 (bases 1 to 181655)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE	JOURNAL	REFERENCE
Homo sapiens chromosome 3, clone RP11-297C5	Unpublished	2 (bases 1 to 181655)

AUTHORS
Britten, B., Linton, L., Nussbaum, C., Landet, E., Aorham, H., Allen, N., Anderson, S., Baldwin, J., Bana, N., Bastien, V., Beda, F., Boguslavsky, L., Boukgalter, B., Brown, A., Burkett, G., Campioano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamaras, R., Landers, T., Lebozsky, J., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanik, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Testfayre, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 330 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 25, 2000 this sequence version replaced gi:5547183. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RN/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>

```
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7850
Center clone name: 297_C_5
----- Summary Statistics
```

Sequencing vector: M13: M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 177627 bases at least 40
 Consensus quality: 179713 bases at least 30
 Consensus quality: 180487 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 180955; sum-of-contigs
 quality coverage: 5.0 in Q20 bases; agarose-fp
 quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a "working draft" sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	406: contig of 406 bp in length
*	407	506: gap of 100 bp
*	507	3313: contig of 2807 bp in length
*	3314	3413: gap of 100 bp
*	3414	7825: contig of 4412 bp in length
*	7826	7925: gap of 100 bp
*	7926	13385: contig of 5460 bp in length
*	13386	13485: gap of 100 bp
*	13486	36364: contig of 22879 bp in length
*	36365	36464: gap of 100 bp
*	36465	66064: contig of 29600 bp in length
*	66065	66164: gap of 100 bp
*	66165	105760: contig of 39596 bp in length
*	105761	105860: gap of 100 bp
*	105861	181655: contig of 75795 bp in length

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-297C5"
/clone_lib="RPC1-11 Human Male BAC"
1..406
msc-feature

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

	a	c	g	t	others
BASE COUNT	59958	31496	32944	56553	704
ORIGIN					

Query Match	12.0%;	Score 40;	DB 71;	Length 181655;
Best Local Similarity	52.4%;	Pred. No. 1.5;		

LOCUS AF002985 206996 bp DNA HTG 05-DEC-2000
DEFINITION Homo sapiens chromosome 11 clone CTD-2542M24 map 11q, WORKING DRAFT
SOURCE 41 unordered pieces.
ACCESSION AF002985
VERSION AF002985.1 GI:11559300
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:CTD-2542M24.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 206996)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 206,996 genomic DNA of 11q
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 206996)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-cho, Tsunumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center
COMMENT
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hnp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft11
Center clone name: CTD-2542M24
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 198395 bases at least Q40
Consensus quality: 201470 bases at least Q30
Consensus quality: 202645 bases at least Q20
Insert size: 202996; sum-of-coverage
Quality coverage: 8.33x in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of
41 contigs. The true order of the pieces is not known and the
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 23752 contig of 23752 bp in length
23853 36403 contig of 12551 bp in length
36504 46758 contig of 10255 bp in length
46859 57498 contig of 10640 bp in length
57599 64554 contig of 6956 bp in length
64655 72935 contig of 8281 bp in length
73036 81394 contig of 8359 bp in length
81495 88502 contig of 7008 bp in length
88603 94093 contig of 5491 bp in length
94194 101258 contig of 7065 bp in length
101359 105787 contig of 4429 bp in length
105888 112640 contig of 6753 bp in length
112741 118936 contig of 6198 bp in length
119039 125607 contig of 6569 bp in length
125708 131167 contig of 5460 bp in length
131268 138727 contig of 7460 bp in length
138828 143241 contig of 4414 bp in length
143342 147475 contig of 4134 bp in length
147476 147575 contig of 100 bp in length
147576 151609 contig of 4034 bp in length
151610 151709 contig of 100 bp in length
151710 156821 contig of 5112 bp in length
156822 156921 contig of 100 bp in length
156922 161062 contig of 4141 bp in length
161063 161162 contig of 100 bp in length
161163 163661 contig of 2499 bp in length
163662 163761 contig of 100 bp in length
163762 166532 contig of 2771 bp in length
166533 166632 contig of 2499 bp in length
166633 170243 contig of 2771 bp in length
170244 170343 contig of 3611 bp in length

170344 174453 contig of 4110 bp in length
174554 177123 contig of 2570 bp in length
177224 179605 contig of 2282 bp in length
179706 181344 contig of 1639 bp in length
181445 184052 contig of 2608 bp in length
184153 186128 contig of 1976 bp in length
186229 188419 contig of 2191 bp in length
188520 190571 contig of 2052 bp in length
190672 193870 contig of 3199 bp in length
193971 196360 contig of 2390 bp in length
196461 198058 contig of 1598 bp in length
198159 199874 contig of 1716 bp in length
199975 201469 contig of 1495 bp in length
201570 203031 contig of 1462 bp in length
203132 204175 contig of 1044 bp in length
204276 205618 contig of 1343 bp in length
205719 206996 contig of 1278 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 23752: contig of 23752 bp in length
23753 23852: gap of 100 bp
23853 36403: contig of 12551 bp in length
36404 36503: gap of 100 bp
36504 46758: contig of 10255 bp in length
46759 46858: gap of 100 bp
46859 57498: contig of 10640 bp in length
57499 57596: gap of 100 bp
57599 64554: contig of 6956 bp in length
64555 64654: gap of 100 bp
64655 72935: contig of 8281 bp in length
72936 73035: gap of 100 bp
73036 81394: contig of 8359 bp in length
81395 81494: gap of 100 bp
81495 88502: contig of 7008 bp in length
88503 88602: gap of 100 bp
88603 94093: contig of 5491 bp in length
94094 94193: gap of 100 bp
94194 101258: contig of 7065 bp in length
101259 101358: gap of 100 bp
101359 105787: contig of 4429 bp in length
105788 105887: gap of 100 bp
105888 112640: contig of 6753 bp in length
112641 112740: gap of 100 bp
112741 118936: contig of 6198 bp in length
118939 119038: gap of 100 bp
119039 125607: contig of 6569 bp in length
125608 125707: gap of 100 bp
125708 131167: contig of 5460 bp in length
131168 131267: gap of 100 bp
131268 138727: contig of 7460 bp in length
138728 138827: gap of 100 bp
138828 143241: contig of 4414 bp in length
143242 143341: gap of 100 bp
143342 147475: contig of 4134 bp in length
147476 147575: gap of 100 bp
147576 151609: contig of 4034 bp in length
151610 151709: gap of 100 bp
151710 156821: contig of 5112 bp in length
156822 156921: gap of 100 bp
156922 161062: contig of 4141 bp in length
161063 161162: gap of 100 bp
161163 163661: contig of 2499 bp in length
163662 163761: gap of 100 bp
163762 166532: contig of 2771 bp in length
166533 166632: gap of 100 bp
166633 170243: contig of 2771 bp in length
170244 170343: gap of 100 bp

```
* 170344 174453: contig of 4110 bp in length
* 174554 174553: gap of 100 bp
* 174554 177123: contig of 2570 bp in length
* 177124 177223: gap of 100 bp
* 177224 179605: contig of 2382 bp in length
* 179606 179705: gap of 100 bp
* 179706 181344: contig of 1639 bp in length
* 181345 181444: gap of 100 bp
* 181445 184052: contig of 2608 bp in length
* 184053 184152: gap of 100 bp
* 184153 186128: contig of 1976 bp in length
* 186129 186228: gap of 100 bp
* 186229 188419: contig of 2191 bp in length
* 188420 188519: gap of 100 bp
* 188520 190571: contig of 2052 bp in length
* 190572 190671: gap of 100 bp
* 190672 193870: contig of 3199 bp in length
* 193871 193970: gap of 100 bp
* 193971 196360: contig of 2380 bp in length
* 196361 196460: gap of 100 bp
* 196461 198058: contig of 1598 bp in length
* 198059 198158: gap of 100 bp
* 198159 199874: contig of 1716 bp in length
* 199875 199974: gap of 100 bp
* 199975 201469: contig of 1485 bp in length
* 201470 201569: gap of 100 bp
* 201570 203031: contig of 1462 bp in length
* 203032 203131: gap of 100 bp
* 203132 204175: contig of 1044 bp in length
* 204176 204275: gap of 100 bp
* 204276 205618: contig of 1343 bp in length
* 205619 205718: gap of 100 bp
* 205719 206996: contig of 1278 bp in length.
```

FEATURES

```
source
1..206996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTD-2542M24"
1..23752
misc_feature
/note="assembly_fragment"
23853..36403
misc_feature
/note="assembly_fragment"
36504..46758
misc_feature
/note="assembly_fragment"
46859..57498
misc_feature
/note="assembly_fragment"
57599..64554
misc_feature
/note="assembly_fragment"
64655..72935
misc_feature
/note="assembly_fragment"
73036..81394
misc_feature
/note="assembly_fragment"
```

Query Match 11.9%; Score 39.6; DB 83; Length 206996;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 208 ttattatttttgggtcattgttcagagagcaagacacatcgtcttttgat 267
DB 194579 TTTTTCCTTTCTTTGTTTCCTTGTTCCTGCGAGAAATCATCTTACATATGATATATTTAGAT 194520
Qy 268 attgggaattacattccgttaaggccttgaatgctgcatatgatatacat 317
DB 194519 AACGTAAATGCATCATCTTAACAGTTTCAGATGGCTGAATTTTACAAT 194470

RESULT 8
LOCUS DG087812 1440 bp DNA INV 19-MAR-1997
DEFINITION Drosophila grimshawi Df-1-like protein (Dg1\df1) gene, partial cds.

ACCESSION U07812
VERSION U07812.1 GI:1890773
KEYWORDS Drosophila grimshawi.
SOURCE Drosophila grimshawi.
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1440)
AUTHORS Edwards,K.A., Nakano,Y., Kurihara,J., Kaneshiro,K. and Yamamoto,D.
TITLE Variable sequence conservation between D. melanogaster and Hawaiian picture-winged Drosophila
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1440)
AUTHORS Edwards,K.A., Nakano,Y., Kurihara,J., Kaneshiro,K. and Yamamoto,D.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1997) CCRT, University of Hawaii, 3050 Malle Way #409, Honolulu, HI 96822, USA
LOCATION/Qualifiers
1..1440
/organism="Drosophila grimshawi"
/strain="G1"
/db_xref="taxon:7222"
/note="Hawaiian picture-winged species group"
join(<297..464,336..>971)
/gene="Dg1\df1"
/product="Df1-like protein"
297..971
/gene="Dg1\df1"
join(<297..464,536..>971)
/gene="Dg1\df1"
/note="df1-like gene; similar to Caenorhabditis elegans mitochondrial carrier protein Df-1, PIR Accession Number S55056"

FEATURES

```
source
1..1440
/organism="Drosophila grimshawi"
/strain="G1"
/db_xref="taxon:7222"
/note="Hawaiian picture-winged species group"
join(<297..464,336..>971)
/gene="Dg1\df1"
/product="Df1-like protein"
297..971
/gene="Dg1\df1"
join(<297..464,536..>971)
/gene="Dg1\df1"
/note="df1-like gene; similar to Caenorhabditis elegans mitochondrial carrier protein Df-1, PIR Accession Number S55056"
/codon_start=1
/product="Df1-like protein"
/protein_id="AAB49889.1"
/db_xref="GI:1890774"
/translation="MYSFSSLSNSANMAVKVENNESATTEKANKPVKSFITGGFGTIC NVLSGHPIDITIKVRLQTMRPPEGPQPMYRGTFPCAAKTINNEGVIKQMSAPLTG VAPIFAMCFAGYALGKRLQOREDEAKLTYSOIFVAGSFGSGLFSPFINAPGRIVKLQ TQGTGGGVKRYNMGMLDCAAKLYKGGKRSVFKSCATMLR"
```

Query Match 11.6%; Score 38.6; DB 6; Length 1440;
Best Local Similarity 61.4%; Pred No. 2.7;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 22 gaacgaagggtccacttggctcttcaaggcgcgatcccggtctctcttgatgctc 81
DB 621 GCACGAGGCGCGTCGTCTCTACAAAGGCATCGCGCTCGCTGACGGCGGTGGCCG 680
Qy 82 ctcttgagcaatgaacttgcagcctatgatgcttgcag 122
DB 681 CCATTTTTCGATGTGCTTTGCCGCGCTATGCGCTGGCAAG 721

RESULT 9
LOCUS AC036223 151560 bp DNA HTG 15-JAN-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-362A12 map 2, WORKING DRAFT
SEQUENCE, 1 ordered pieces.
ACCESSION AC036223
VERSION AC036223.3 GI:12229544
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 151560)
Britten,B., Linton,L., Nussbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 2, clone RP11-362A12

Unpublished
2 (bases 1 to 151560)

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Menus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanal,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 2001 this sequence version replaced gi:7770602.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L9424

Center clone name: 362_A_12
Summary Statistics
Sequencing vector: M13; M7815; 48% of reads
Sequencing vector: Plasmid; n/a; 52% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151269 bases at least Q40
Consensus quality: 151406 bases at least Q30
Consensus quality: 151491 bases at least Q20
Insert size: 15000; agarose-fp
Insert size: 151560; sum-of-ctnigs
Quality coverage: 10.8 in Q20 bases; agarose-fp
Quality coverage: 10.7 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 151560: contig of 151560 bp in length.
assembly_fragment.

FEATURES
SOURCE

Location/Qualifiers
1. 151560
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

BASE COUNT
ORIGIN
44866 a 27588 c 28440 g 50665 t 1 others

Query Match 11.6% Score 38.6; DB 72; Length 151560;
Best Local Similarity 55.6% Pred. No. 3.7; Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Db 132 ataaagatggaataatccagatgacacacagagatcttcacgaagaacagatga 191
54111 ACAGAAATGCAAAATTCCTTATATATCTTCCTCAGGTGGGATGATCATGATCA 54170
Db 192 tcaacacatgcaactttatattatatttttgggtccatgctgacgaagaacaa 251
54171 TCTTCTTCACAAATTCCTTATATATCTTCCTCAGGTGGGATGATCATGATCA 54230
Qy 252 tctgctctttttt 264
Db 54231 TATTTGTTCTTGT 54243

RESULT 10
AC073167/c 174217 bp DNA HTG 03-FEB-2001
LOCUS Homo sapiens chromosome 15 clone RP11-31766 map 15, WORKING DRAFT
DEFINITION AC073167
ACCESSION AC073167
VERSION AC073167.2 GI:12658040
KEYWORDS HTG; HTGS; PHASEL; HTGS; DRAFT.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174217)
AUTHORS Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone RP11-31766
COMMENT Unpublished
2 (bases 1 to 174217)

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 174217)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Menus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanal,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (09-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 3, 2001 this sequence version replaced gi:8389488.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L10035
Center clone name: 317_G_6
Sequencing vector: M13; M7815; 58% of reads

```

Sequencing vector: Plasmid; n/a; 42% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 170367 bases at least Q40
Consensus quality: 172109 bases at least Q30
Consensus quality: 172716 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 173217; sum-of-contigs
Quality coverage: 6.7 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 b.
NOTE: This is a 'working draft' sequence. It currently
consists of 11 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 60103: contig of 60103 bp in length
60104 60203: gap of 100 bp
60204 62204: contig of 2001 bp in length
62205 62304: gap of 100 bp
62305 66315: contig of 4011 bp in length
66316 66415: gap of 100 bp
66416 71015: contig of 4600 bp in length
71016 71115: gap of 100 bp
71116 80999: contig of 9884 bp in length
81000 81099: gap of 100 bp
81100 93750: contig of 12651 bp in length
93751 93850: gap of 100 bp
93851 109484: contig of 15634 bp in length
109485 109584: gap of 100 bp
109585 126301: contig of 16717 bp in length
126302 126401: gap of 100 bp
126402 144194: contig of 17793 bp in length
144195 144294: gap of 100 bp
144295 172644: contig of 28350 bp in length
172645 172744: gap of 100 bp
172745 174217: contig of 1473 bp in length.
Location/Qualifiers
1..174217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-317G6"
/clone_lib="RPC1-11 Human Male BAC"
1..60103
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
60204..62204
/note="assembly-fragment"
62305..66315
/note="assembly-fragment"
66416..71015
/note="assembly-fragment"
71116..80999
/note="assembly-fragment"
81100..93750
/note="assembly-fragment"
93851..109484
/note="assembly-fragment"
109585..126301
/note="assembly-fragment"
126402..144194
/note="assembly-fragment"
144295..172644
/note="assembly-fragment"
172745..174217
/note="assembly-fragment"
clone_end:r7
vector_side:right"

```

```

BASE COUNT      51662 a 32760 c 33482 g 55309 t 1004 others
ORIGIN
Query Match      11.6%; Score 38.6; DB 75; Length 174217;
Best Local Similarity 52.1%; Pred. No. 3.7;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 120 aagaagcagatgataaagaatgagaaatccagagtgacacagagagatctgcga 179
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 59179 AAGAAATCAATTAATTCAGACAGTAATAATGGCAGAAATTAATATGATAGCTTGAGA 59120
QY 180 aagaagcagtgatcaacacatgcacttttatttatttcttggtgcattgtgcagaga 239
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 59119 AGATCAATCAATTAATTCATGAGAACTTCATCCCTTGTTGATGATTAACAACCTAAGACT 59060
QY 240 agcaacagacatctgccttttttgatatttgggaattcattcc 284
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 59059 AACGACAGAACATCTGTTTAATACGATATTTTACGACGATTTCC 59015

RESULT 11
AC020137      33714 bp      DNA      HTG      03-JAN-2000
AC020137
LOCUS
DEFINITION
ACCESSION
AC020137.1 GI:6664760
VERSION
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 33714)
AUTHORS
Adams,M. and Venter,J.C.
TITLE
Direct Submission
JOURNAL
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10212166 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..33714
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT      9941 a 7271 c 7102 g 9400 t
ORIGIN
Query Match      11.5%; Score 38.4; DB 66; Length 33714;
Best Local Similarity 54.2%; Pred. No. 3.8;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 120 aagaagcagatgataaagaatgagaaatccagagtgacacagagagatctgcga 179
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 21218 AAGAAATCAATTAATTCAGACAGTAATAATGGCAGAAATTAATATGATAGCTTGAGA 21277
QY 180 aagaagcagtgatcaacacatgcacttttatttatttcttggtgcattgtgcagaga 239
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 21278 AGAAGTCAGTACTTCCCTTTAAATTTTAATTAATTAATTAATTCCTTTAACAANA 21337
QY 240 agcaacagacatctgccttttt 263
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 21338 AATATAAAGTAATCAATATTTT 21361

RESULT 12
AE003647/c
LOCUS      AE003647      262205 bp      DNA      INV      04-OCT-2000

```

DEFINITION	Drosophila melanogaster genomic scaffold 142000013386055 section 40 of 63, complete sequence.
ACCESSION	AE003647 AE002690
VERSION	AE003647.1 GI:7298216
KEYWORDS	HTG.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 262205) Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blasej,R.G., Champagne,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abrell,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Bereman,B.P., Bhandari,D., Bolintsov,S., Botkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotilier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evansgelist,C.C., Ferrara,C., Ferrieria,S., Fleischmann,W., Foster,C., Gabrielian,A.E., Gary,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kienlen-Darmon,M., Kim,C., Kimbrell,M., Kirkpatrick,D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,D., Li,Z., Liang,Y., Lin,X., Liu,X., Maizel,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Modyaty,C., Morris,J.P., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusser,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Sanders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Sytkas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Ye,Q.A., Ye,J., Yen,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
TITLE	The genome sequence of Drosophila melanogaster
JOURNAL	Science 287 (5461), 2185-2195 (2000)
MEDLINE	20196006
REFERENCE	2 (bases 1 to 262205)
AUTHORS	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE	Direct Submission
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
FEATURES	location/Qualifiers
source	1..262205 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="2L" complement(join(716..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5822,6652..6679)) /gene="gift" /product="CT40006" /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" complement(join(716..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5822,6513..6676)) /gene="gift"
gene	1..262205 /product="CT35206" /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" complement(join(1053..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5643)) /gene="gift" /note="gift gene product [alt 2]" /codon_start=1 /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" /protein_id="AAF53451.1" /translation="MTLRGNPPKKECKMIRAFPSAMDEKYETITASIKNAIOEOK KNSGISEFQRLRNATNVLHKGNNLYGLREVSVSEHLEHVRADVLAISNPLPK LNQAWTDHQTSMVMDILMYDRYVQOREVDVNYNLGLIFRQVYVSEIQAKR EKILGVMEEHGEALINHLAIKNACSMITLIGINSTVYEEDPEKFLAQSAAFKFE SONFLAENNAGVYIKKEARITEESSRALYLDKDEPRIVAVEEELIKKMRPIVE MENSQVVMYIKNSKTEDLACTYKLFESRLKEEGKIYADMSAYLRQGMIVKEEENG NNPIITFVONLIDKDRPOPLVHSPANDIRKNVYSSDEPHNLNNKSPYLSLFI DNLKKGKGSQDETESLIDRTWVIFRLLEKDVPERYKTHLAKRLILNKSYSDF EKNMISKLKTECCQFTSKLEGFKMSVSNITIMEFKNFVNNNLSLGLVETVAIL TTGFWPTQATPNCNIPAPREAFDIFKNFVYLNKHSRGRLTLQPOMGTAVINAVEGR KAVESKDKDADSSSSGCGVPTTKRHLIYQSTOMCYLLFFNNNDVLYTDIOET DIPERLVVALOISLGMKPAORLIVNSKTRKIDEPTEPFVNDAFNSKPRVKOT VVAKGSPEPERKTRGKVEDDRKHEEAIVIMKARKRLANLNLVSDVTSQLSKSEL PSVPIFKRIIEGIEREYIQRSPEDKRYNYIA"
CDS	complement(join(1053..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5643)) /gene="gift" /note="gift gene product [alt 1]" /codon_start=1 /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" /protein_id="AAF53450.1" /translation="MTLRGNPPKKECKMIRAFPSAMDEKYETITASIKNAIOEOK KNSGISEFQRLRNATNVLHKGNNLYGLREVSVSEHLEHVRADVLAISNPLPK LNQAWTDHQTSMVMDILMYDRYVQOREVDVNYNLGLIFRQVYVSEIQAKR EKILGVMEEHGEALINHLAIKNACSMITLIGINSTVYEEDPEKFLAQSAAFKFE SONFLAENNAGVYIKKEARITEESSRALYLDKDEPRIVAVEEELIKKMRPIVE MENSQVVMYIKNSKTEDLACTYKLFESRLKEEGKIYADMSAYLRQGMIVKEEENG NNPIITFVONLIDKDRPOPLVHSPANDIRKNVYSSDEPHNLNNKSPYLSLFI DNLKKGKGSQDETESLIDRTWVIFRLLEKDVPERYKTHLAKRLILNKSYSDF EKNMISKLKTECCQFTSKLEGFKMSVSNITIMEFKNFVNNNLSLGLVETVAIL TTGFWPTQATPNCNIPAPREAFDIFKNFVYLNKHSRGRLTLQPOMGTAVINAVEGR KAVESKDKDADSSSSGCGVPTTKRHLIYQSTOMCYLLFFNNNDVLYTDIOET DIPERLVVALOISLGMKPAORLIVNSKTRKIDEPTEPFVNDAFNSKPRVKOT VVAKGSPEPERKTRGKVEDDRKHEEAIVIMKARKRLANLNLVSDVTSQLSKSEL PSVPIFKRIIEGIEREYIQRSPEDKRYNYIA"
gene	1..262205 /product="CT35205" /db_xref="FLYBASE:FBan0015261" /db_xref="FLYBASE:FBgn0028510" complement(join(8905..9707)) /gene="Bg:DS07851.3" /note="CG15261" /map="35C4-35C4" /db_xref="FLYBASE:FBan0015261" /db_xref="FLYBASE:FBgn0028510" complement(join(8948..9069,9299..9528,9601..9665)) /note="Bg:DS07851.3 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0015261"
mRNA	1..262205 /product="CT35205" /db_xref="FLYBASE:FBan0015261" /db_xref="FLYBASE:FBgn0028510" complement(join(8905..9707)) /gene="Bg:DS07851.3" /note="CG15261" /map="35C4-35C4" /db_xref="FLYBASE:FBan0015261" /db_xref="FLYBASE:FBgn0028510" complement(join(8948..9069,9299..9528,9601..9665)) /note="Bg:DS07851.3 gene product" /codon_start=1 /db_xref="FLYBASE:FBan0015261"
mRNA	1..262205 /product="CT35205" /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" complement(join(716..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5822,6513..6676)) /gene="gift"
mRNA	1..262205 /product="CT35205" /db_xref="FLYBASE:FBan0011861" /db_xref="FLYBASE:FBgn0001980" complement(join(716..1205,1275..1413,1475..1698,1824..1966,2032..2146,2209..2379,2442..2612,2675..3436,3500..3883,5584..5822,6513..6676)) /gene="gift"

```

/db_xref="FLYBASE:FBgn0028510"
/protein_id="AAFS3452.1"
/db_xref="GI:7298219"
/translacion="MSTIVRKLIISTANAAKFAVAFYNOAVADRTVYVSGIGLIDKDTM
KIVGSPTEQOKALENLEAVLKAADSGVDVINYNTVFLKDLNDFGAVNEVYKKRVPNNK
DEPFRSCFOVAKLMDALVEIECIALGSVETKIVQ"
<19706..>20584
/gene="BG:DS07851.4"
/product="CT35209"
/db_xref="FLYBASE:FBan0015263"
/db_xref="FLYBASE:FBgn0028853"
/evidence=not_experimental
<19706..>20584
/gene="BG:DS07851.4"
/map="35C4-35C5"
/note="CG15263"
/db_xref="FLYBASE:FBan0015263"
/db_xref="FLYBASE:FBgn0028853"
/evidence=not_experimental
19706..20584
/gene="BG:DS07851.4"
/note="BG:DS07851.4 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015263"
/db_xref="FLYBASE:FBgn0028853"
/evidence=not_experimental
/protein_id="AAFS3453.1"
/db_xref="GI:7298220"
/translacion="MAAFKYSIAFLVILVACGVTVVPAKHPRSROEAENMSKFSVLQK
EQLSAGSPRRNDVUGNARAIYVYODKGFITGRGTFQMTSPNPNNDGYINPQFPADE
SSGIGIGIONDKSSKFTVLFSTWDSAPFEDDADACFTGGEGVGYKCYNSTPFLISN
AIYLVDAIEBGSFTRGYISDPQENIDSVTSQVNRHLIGELITPHEKLTNLIIPGYNN
ENYDRDTRGAEAFELVYVNCAPHQYTKGHANTKFTBAYILNSECEVENIILVLTADRK
KSTLYTRPDLNTYN"
complement(join(25368..25523,26175..>26872))
/gene="BG:DS07851.8"
/product="CT35204"
/db_xref="FLYBASE:FBan0015260"
/db_xref="FLYBASE:FBgn0028850"
complement(join(25368..>26872))
/gene="BG:DS07851.8"
/note="CG15260"
/map="35C4-35C5"
/db_xref="FLYBASE:FBan0015260"
/db_xref="FLYBASE:FBgn0028850"
complement(join(25368..25523,26175..26807))
/gene="BG:DS07851.8"
/note="BG:DS07851.8 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015260"
/db_xref="FLYBASE:FBgn0028850"
/protein_id="AAFS3454.1"
/db_xref="GI:7298221"

Query Match 11.5%; Score 38.4; DB 4; Length 262205;
Best Local Similarity 54.2%; Pred. NO. 4.4;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

CDS
120 aagaagcgcaatgatataaagatagagaaataatccatgagtgatcaacacgagagatcttgcca 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6993 AAGAAATCAATATATATAGTGCACATATATCTTCTGTGTTCTGCTTGATGCAAGCCTACTA 6934

QY 180 aagaagcagtgatcaacatgcaacttttatttattatttttggtgcatggttcagagaa 239
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6993 AGAATCTAGTACTTCTCTTTAATTTTATATTTATTTAATATTTATATTTCTTTACCAATA 6874

QY 240 agcaacagacacatctgccttttt 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6873 AAATTAAGTAAGTACTATATTTT 6850

RESULT 13
PROSADH06/c

```

LOCUS	DROSADH06	316756 bp	DNA	INV	21-MAR-2000
DEFINITION	Drosophila melanogaster, chromosome 2L, region 34C4-36A7 (Adh region), section 6 of 10 of the complete sequence.				
ACCESSION	AE003412 AC002437 AC001274 AC002502 AC000658 AC000335 AC001801 AC001810 AC001798 AC001799 AC000316 AC000661 AC000665 AC000662 AC000665 AC000659 AC000660 AC000657 AC000664 AC000656 AC000668 AC000669 AC000670 AC000671 AC000672 AC000673 AC000674 AC000675 AC000676 AC001804 AC001805 AC000677 AC001803 AC0002875 AC000675 AC0002435 AC001239 AC000447 AC001668 AC001240 AC001123 AC000448 AC000242 AC000256 AC001243 AC000449 AC001669 AC001244 AC000450 AC000451 AC001670 AC000453 AC000454 AC000455 AC001245 AC001246 AC001247 AC000456 AC001671 AC001672 AC000457 AC001248 AC001249 AC001673 AC000458 AC001250 AC001674 AC000459 AC001251 AC001675 AC001252 AC000460 AC000461 AC000462 AC001676 AC0002517 L77028 L77020 L77021 L79887 L77022 L77025 L77027 L79888 AC000128 L81343 L77027 L77024 L81346 L79886 AC000655 L77029 L77026 L81344 L81345 AE003406				
VERSION	AE003412.1 GI:7287753				
KEYWORDS	HTG.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 316756) Ashburner,M., Miska,S., Roote,J., Lewis,S.E., Blazer,R., Davis,T., Doyle,C., Galle,R., George,R., Harris,N., Hartzell,G., Harey,D., Hong,L., Houston,K., Hoskins,R., Johnson,G., Martin,C., Moshrefi,A., Palazzolo,M., Reese,M.G., Spradling,A., Tsang,G., Wan,K., Whitelaw,K., Celniker,S. and Rubin,G.M. An exploration of the sequence of a 2.9-Mb region of the genome of Drosophila melanogaster: the Adh region Genetics 153 (1), 179-219 (1999)				
TITLE	2 (bases 1 to 316756) Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazer,R.G., Buehnhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummstl,S.R., Katra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacble,J.M., Park,S., Pfeiffer,B., Poon,L., Sequiera,A., Sethi,H., Snir,E., Svaykas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.				
JOURNAL MEDLINE	Direct Submission Submitted (08-MAR-2000) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, MS 64-121, Berkeley, CA 94720, USA				
PUBMED	On Mar 22, 2000 this sequence version replaced gi:2708244 gi:3046226 gi:2335082 gi:3168611.				
AUTHORS	Submitted by the Berkeley Drosophila Genome Project. For more information, visit the BDGP Web site: http://www.fruitfly.org/ . This is the finished sequence of 34C4-36A7. The orientation of this sequence along the chromosome is left to right. This sequence was annotated along Sma Mspa (smaefrutfly.berkeley.edu) on behalf of the Berkeley Drosophila Genome Center. Coding sequences are predicted based on computational analysis, using both gene and CDS prediction programs and matches to other sequences. These predictions and matches have been evaluated by the annotators and may have been refined by hand. The annotators have also used their judgement about which matches to include in this record. The annotations on this sequence can be examined in more detail from http://www.fruitfly.org/publications/Adh.html The annotation syntax used in this record is documented at				
REFERENCE					
COMMENT					

ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation_README.v
1.2.

FEATURES

source

```
location/Qualifiers
1.316756
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="34C4-36A7"
/clone="P1 DS01368 (D100), P1 DS08249 (D25), P1 DS08284
(D247), P1 DS00941 (D28), P1 DS08220 (D39), P1 DS00180
(D29), P1 DS01514 (D33), P1 DS00131 (D35), P1 DS05899
(D22), P1 DS01759 (D41), P1 DS01523 (D34), P1 DS01652
(D43), P1 DS03792 (D31), P1 DS01068 (D37), P1 DS06238
(D26), P1 DS08340 (D38), P1 DS04641 (D30), P1 DS01160
(D42), P1 DS01486 (D32), P1 DS09219 (D101), P1 DS07721
(D20), P1 DS00810 (D21), P1 DS06874 (D19), P1 DS04311
(D102), P1 DS03144 (D106), P1 DS03323 (D127), P1 DS01219
(D129), P1 DS00929 (D48), P1 DS04929 (D104), P1 DS03192
(D103), P1 DS09194 (D267), P1 DS07295 (D47), P1 DS05639
(D111), P1 DS07851 (D49), P1 DS03023 (D8), P1 DS01845
(D7), P1 DS04862 (D107), P1 DS07108 (D51), P1 DS09217
(D109), P1 DS02252 (D97), P1 DS00365 (D108), P1 DS07486
(D46), P1 DS00913 (D24), P1 DS04095 (D112), P1 DS02795
(D50), P1 DS07473 (D113), P1 DS02740 (D27), P1 DS09218
(D114), P1 DS02780 (D23), BAC BACR4B02 (D482), BAC
BACR44L22 (D483)"
/clone_lib="P1 library, partial Sau3a in pMS582tet14ad10
and Rpt1-98 (Roswell Park Cancer Institute Drosophila
melanogaster BAC library, partial EcoRI in pBACE3.6)"
complement(join(921..1425,1495..1615,1684..1830,
1893..2458,5518..3304,3366..4816,4873..7267,7332..7532,
7599..7948,8279..8539,9219..9322,13473..13614))
/gene="ck"
/complement(921..13614)
/gene="ck"
/complement(join(1219..1425,1495..1615,1684..1830,
1893..2458,5518..3304,3366..4816,4873..7267,7332..7532,
7599..7948,8279..8539,9219..9322))
/gene="ck"
/feature="ck"
/symbol="ck; synonym="BG:DS00929.11;
match-method:"sim4", score:"1000.0", desc:"Drosophila
melanogaster crinkled (ck) mRNA, sequence from D. Kiehardt
(personal communication)", match-method:"BLASTX",
version:"2.0a19mp-wasnu [05-Feb-1998] [build sol2.5-ulttra
01:47:30 05-Feb-1998]", score:"2202.0",
desc:"trembl:Q13402:MYOSIN VIIA. organism:HOMO SAPIENS
(HUMAN). dbxref:genbank: U39226; g1235670; -",
species:"HOMO SAPIENS"
/codon_start=1
/evidence="experimental"
/protein_id="AA04915.1"
/db_xref="GI:7287877"
/db_xref="FLYBASE:FBgn0000317"
/translation="MTYVTRGDIYMIIEPAGSGREDDVAICARVSAECRRIOVRDDCD
EVLADERRIKAMHASSVQGVEDMTISGLDHEGILNLIRKENLITYTGSILVA
VNPYOLPIYTGDIKIKRKRIKIGELPPHFAIGDNAYAMKRRQOCVITSESA
KTESTKLILOYLAISGKSHWEOQLLENPALEAGNKKTINDOSSRFKIDIH
FSANGVEGAKRBOYILEKRSIVSONSENHYVCIILAGLSDKSRIDIGMAAYT
KYLGNSTICEGRDDAEERDIISAKVLLFSQOETWEIKLILAHGNTKRYAT
VVDNLDKTEIPEHINERVAAGLGLPIOLPILDIRTRTLRAHGETVYSTSRDSDVY
RDAFVKGIYGRMFVHVRKINTALFKPRGSRNAIGVLIFFGFENFQNSPEFCIN
ANENLQGFVQHIFLEQDEYNHEAIMMHOIEFVNDQDADLDAIKDNLIMALIDEBA
RFPKQDOTMLAKLHTKSHGSHKYLKPSIINSFGNHNGAVFVYPTRELDKNRPT
ESPDLHLVSGSTNKKRLROIADODIEGASERKRPPTLSHQFSLDALAKTTSSCOO
PFRICIKRPNELKAPMFDRLCCQRLYSQSMETIRIRRAGYPIRHRFRFVEYRRL
IPGVPAHRDTCOATSIRICAVVAGKSDYDGLSHKVRLLKAHDLFLQEDRVLTRKI
LILQRSIRGWYRRRLRLRAAATVQRFKQGAQRKRRMYRGMRLQALIRSRL
SHRFRLRHGIIVGLQAHARGYLVRREYGMMAVVIKQSHVRRMYRIRKRLLEK
QFAEVLQRLKEDELBHGRNKHAREIAEDHVRJOLELERREIOLENNRRREYV
NITINDARKOEPEVDGKLEAMFDELPSSSDAPPHGGRSTSVFNDLPHAOVND
DIAPHIHISEDLEDFKFOKPAATYFQGNVNHQYAKKALKHLPLPLHTOGDGLAAO
```

gene

CDS

```
complement(join(15000..16900,16927..17232))
/gene="Tf1ts"
/complement(15000..17232)
/gene="Tf1ts"
/complement(15296..16087)
/gene="Tf1ts"
/feature="Tf1ts; synonym="BG:DS00929.12;
match-method:"sim4", score:"990.0",
desc:"Genbank:126091:Drosophila melanogaster
transcription elongation factor (Tf1ts) gene, complete
cds. CDS:1124..1915; PID:9418355", species:"Drosophila
melanogaster", match-method:"BLASTX", [build sol2.5-ulttra
version:"2.0a19mp-wasnu [05-Feb-1998] [build sol2.5-ulttra
01:47:30 05-Feb-1998]", score:"754.0",
desc:"SwissProt:P10712:TRANSCRIPTION ELONGATION FACTOR
S-II (TRANSCRIPTION ELONGATION FACTOR A). dbxref:EMBL:
M18209; g201937; -; EMBL: M18210; g201939; -. PIR: A29950;
A29950; HSP: P23193; LTF: MKD; MGI:1073368; TCEAR
PROSITE; PS00466; Tf1ts 1.", species:"MUS MUSCULUS"
/codon_start=1
/evidence="experimental"
/protein_id="AA04916.1"
/db_xref="GI:7287878"
/db_xref="FLYBASE:FBgn0010422"
/translation="MTVNELRKSSKDEVTALAKTLIKNNKRELPASPPTPNNSAK
EGSSNNSASKSTSAKSSSSISGCKSSSSSSSKDEKKGSTSSSOTSPSGMTDA
VRICKREMLATALKIGEVPEGCEPEMAALELDEAIYSEFNNTDMKNIIRSVANTL
KDPKNPGLRGNEFMCGAVTAKOLAKMKEPEEMASDEMKKLRKFKYEAINDAQLAVQGT
KTDLKCAKCKKRCCTYNLOTRSDAPMTTFVWCNCGRMKRF"
join(17510..17567,24187..24234,24288..24675,28143..28432,
28492..29235,29294..29507,29572..29843,29915..30116)
/gene="vas"
17510..30116
Query Match 11.5% Score 38.4; DB 6; Length 316756;
Best Local Similarity 54.2% Pred. No. 4.4;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
Qy 120 aagaaggaagatgaanaagatgagagaanaatcagtgagtaacacagagagatcttgcca 179
Db 228120 AAGAAATCAATAAATATTAGTACAAATATATCTTGCTTGCTTGATGACGAAGCTACTA 228061
Qy 180 agaaagcagtgatcaacatgcacacttttattattatttttgggtccatgltgcagaga 239
Db 228060 AGAAGTCAGTACTTCTCTTTAATTTTAAATTTTAAATTTTAAATTCCTTTAACAAATA 228001
Qy 240 agcaacagacatcgcctttt 263
Db 228000 AAATTAAGTAAGTACATATATTTT 227977
RESULT 14
```


U32813 LOCUS 11681 bp DNA BCT 29-MAY-1998
DEFINITION Haemophilus influenzae Rd section 128 of 163 of the complete genome.
ACCESSION U32813 L42023
VERSION U32813.1 GI:1574796
KEYWORDS
SOURCE
ORGANISM Haemophilus influenzae Rd.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.
REFERENCE 1 (bases 1 to 11681)
AUTHORS Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirsnes, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G.G., Fitzhugh, W., Fields, C.A., Gocayne, J.D., Scott, J.D., Shiley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudel, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhmann, J.L., Geoghagen, N.S., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C.
TITLE Whole-genome random sequencing and assembly of Haemophilus influenzae Rd
JOURNAL Science 269 (5223), 496-512 (1995)
MEDLINE 9530630
REFERENCE 2 (bases 1 to 11681)
AUTHORS Tatusov, R.L., Muthaegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Borodovsky, M., Rudd, K.E. and Koonin, E.V.
TITLE Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli
JOURNAL Curr. Biol. 6 (3), 279-291 (1996)
MEDLINE 96398784
REFERENCE 3 (bases 1 to 11681)
AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 4 (bases 1 to 11681)
AUTHORS White, O., Clayton, R.A., Kerlavage, A.R. and Fleischmann, R.D.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REMARK The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes
5 (bases 1 to 11681)
REFERENCE 5 (bases 1 to 11681)
AUTHORS White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D., Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REMARK The whole genome was shifted by 588 nucleotides for a new start on Oct 1, 1996 this sequence version replaced g1:1222102.
COMMENT location/Qualifiers
FEATURES
source 1..11681
/organism="Haemophilus influenzae Rd"
/db_xref="taxon:71421"
gene 86..913
/gene="HI1336"
CDS 86..913
/gene="HI1336"
/note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776 GB:X68777 percent identity: 54.15; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydropyrimidine synthase (folP-1)"
/protein_id="AAC22983.1"
/db_xref="gi:1574797"
/translation="MKLVANNKCLDSVQINGILNFTPDPSDSGQFFSLDKALFOV

EMKLEEGATIIDIGCESTRPNADVESEDEELHRRVVPVVEAVNPRDCAISVSSKAVY
KREANVGNMDLINDIRALOEPRALETAKKALPVCIMHMQOPRTMAYENYVQD
VLAFLOKRTNECLSGVIRKENLIMWGFSGFSGVSNYHOLQNLNFEQOICYPVLAG
SRKSMIGAVLDKPYDORIIIGSAGALIAVOKAKILRVHDAVAAATSDMLKVMQATANA"
951..2288
/gene="HI1337"
951..2288
/gene="HI1337"
/note="similar to GB:L12968 SP:P31120 GB:U01376 PID:290443
PID:606114 percent identity: 73.26; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="mrSA protein (mrSA)"
/protein_id="AAC22984.1"
/db_xref="gi:1574798"
/translation="MANRRYFCTDGVRCGVAYPIPTPDLALGLMAAGKVLASQSKM
VLIGDRTISGCMESLEAGLAAGLSAATLGPMPALIVITPTPFAEGIVTAS
HNPIYDNDIKFPSAKGTLPDPIEATITAMLEQPMDCVSAELGKASRINDAAGRTIE
FCKGTPAHNLGEGTKIVDCNAGVYTHIAPVNLKELAEYIEIGDPNGININKEGL
ANDVVALQAKVETKADVGLAVDGGDITMVDHLGNKVVDDQILFIARALRSGQL
KGVGTGIVASLAVLAAMAQHKLSINELSAVKLPPQVILNVRFGGNGEPLSDAVKS
VAAEVEKRLGEGRIILRRSGTEPLIRVWVEQDAELAQCAEEIAEAVKIN"
2347..2841
/gene="HI1338"
2347..2841
/gene="HI1338"
/note="similar to GB:U00096 PID:1788681 percent identity:
40.26; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22985.1"
/db_xref="gi:1574807"
/translation="MNIFLMRGEAEVANSRKHLLTVYSGKAFLOGMLKHLST
LVINSLDRIVSPYRAQETPEHVNQADPLEENKEETIEGTTPGHAHSVDTYELV
KDEGKSVLYSHLPVGEIYALVETLGRKNPISFYATTIQLLMQNKSEILMRQASV
ITLK"
3043..3432
/gene="HI1339"
3043..3432
/gene="HI1339"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/transl_table=11
/product="H. influenzae predicted coding region HI1339"
/protein_id="AAC22992.1"
/db_xref="gi:1574799"
/translation="MEKIMKKUTLLALVLGSLAVVTGCFDKQKAKQVEDTQTVASVA
SETKDAANTMTVEKKAQQLSTDVKNFVAEKVEDAKFVIRKATEAASEKGEKMEA
SEKASEMEKAEVSEKATQAVDAVKEATK"
complement(3510..4835)
/gene="HI1340"
complement(3510..4835)
/gene="HI1340"
/note="similar to PID:882689 GB:U00096 SP:Q46920
PID:1789158 percent identity: 58.08; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22986.1"
/db_xref="gi:1574800"
/translation="MNYODNSLSKSLGKTEYASQYDRTLLOPVPLANRDLGAGITQ
NOPETIGADIMTAYEISMLNEKGLPOVALIADYIDYQSONLIESKSFVLNYSNQS
FTDENAOOTMOROLSECAQGDYKRLRPAVAYDAQKRNHLOQCIDQDIEITSYER
NANMLKDCVSEIYEKLVSHLKSNCILITNQPDMGTQDAQAILIARNNKLNEFTOR
KTAEGTILRNLLNKLKNEALNITFPIHINVKTAGVNLNVPVSVIANRPVKAQGRLOS
AFKNKAKQKSWFPEVNLGASLSASTVGTALNHPVAGVIGLSLPLNNTYKMYN
KISEADYETARNEFORITFTALNNVDITNVPAPFTQASQSLSNLQOQHSNQTITQYRN
RINAGVSELREVLVAANREKSSQLAILNAKQOYLQSENAVYSSMAGTYL"

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:01 ; Search time 257.37 Seconds
(without alignments)
755.330 Million cell updates/sec

Title: US-09-619-643-5

Perfect score: 333
Sequence: 1 tgaatacttgcacatcctt.....acattacatgagaagtgc 333

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: N_Geneseq_0401:*
2: /SID52/gcgcdata/geneseq/geneseqn/NA1980.DAT:*
3: /SID52/gcgcdata/geneseq/geneseqn/NA1981.DAT:*
4: /SID52/gcgcdata/geneseq/geneseqn/NA1982.DAT:*
5: /SID52/gcgcdata/geneseq/geneseqn/NA1983.DAT:*
6: /SID52/gcgcdata/geneseq/geneseqn/NA1984.DAT:*
7: /SID52/gcgcdata/geneseq/geneseqn/NA1985.DAT:*
8: /SID52/gcgcdata/geneseq/geneseqn/NA1986.DAT:*
9: /SID52/gcgcdata/geneseq/geneseqn/NA1987.DAT:*
10: /SID52/gcgcdata/geneseq/geneseqn/NA1988.DAT:*
11: /SID52/gcgcdata/geneseq/geneseqn/NA1989.DAT:*
12: /SID52/gcgcdata/geneseq/geneseqn/NA1990.DAT:*
13: /SID52/gcgcdata/geneseq/geneseqn/NA1991.DAT:*
14: /SID52/gcgcdata/geneseq/geneseqn/NA1992.DAT:*
15: /SID52/gcgcdata/geneseq/geneseqn/NA1993.DAT:*
16: /SID52/gcgcdata/geneseq/geneseqn/NA1994.DAT:*
17: /SID52/gcgcdata/geneseq/geneseqn/NA1995.DAT:*
18: /SID52/gcgcdata/geneseq/geneseqn/NA1996.DAT:*
19: /SID52/gcgcdata/geneseq/geneseqn/NA1997.DAT:*
20: /SID52/gcgcdata/geneseq/geneseqn/NA1998.DAT:*
21: /SID52/gcgcdata/geneseq/geneseqn/NA1999.DAT:*
22: /SID52/gcgcdata/geneseq/geneseqn/NA2000.DAT:*
23: /SID52/gcgcdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.2	11.5	1830121	17 T42063	Haemophilus influe
2	37	11.1	1443	20 X25344	Human adenine nucl
3	37	11.1	2095	20 X25343	Human adenine nucl
4	37	11.1	2876	21 C90467	Human uncoupling p
5	37	11.1	3145	21 C77456	Human ORFX ORF3011
6	36.4	10.9	1267	20 Z20025	Wheat brittle-1 pa
7	34.8	10.5	2750	14 Q38808	G6PD coding sequen
8	33.8	10.2	32768	20 X13037	Enterococcus faeca
9	33.6	10.1	1172	21 F09810	Fusarium venenatum
10	33.4	10.0	821	21 A02473	Human colon cancer
11	33	9.9	1195	21 C49878	Arabidopsis thalia

12	33	9.9	1198	21 C33296	Arabidopsis thalia
13	33	9.9	3772	20 X81263	Altermonas lipase
14	32.8	9.8	95223	21 F22282	BAC containing rep
15	32.2	9.7	1289	21 C42524	Arabidopsis thalia
16	32	9.6	255	19 X11141	Human diallelic po
17	32	9.6	1344	21 C90456	Human uncoupling p
18	32	9.6	1549	22 F27726	Human transport pr
19	31.8	9.5	829	21 C49805	Arabidopsis thalia
20	31.8	9.5	1331	21 C49806	Arabidopsis thalia
21	31.8	9.5	1334	21 C35455	Arabidopsis thalia
22	31.8	9.5	1808	22 C99806	Skin cell cDNA, SE
23	31.8	9.5	1816	21 Z61789	cDNA encoding murti
24	31.6	9.5	1816	22 C99722	Skin cell cDNA, SE
25	31.6	9.5	159	21 C54856	Arabidopsis thalia
26	31.6	9.5	1243	19 T98697	DNA encoding a S.
27	31.6	9.5	1244	21 C38049	Arabidopsis thalia
28	31.6	9.5	1246	21 C51062	Arabidopsis thalia
29	31.6	9.5	1249	21 C33981	Arabidopsis thalia
30	31.6	9.5	1249	21 C51061	Arabidopsis thalia
31	31.6	9.5	11340	19 V52280	Streptococcus pneu
32	31	9.3	691	20 X30338	DNA encoding a hum
33	30.8	9.2	1013	21 A46319	Exon 11 of Interph
34	30.8	9.2	1145	20 X13367	Enterococcus faeca
35	30.8	9.2	1853	21 C41782	Arabidopsis thalia
36	30.8	9.2	3383	17 T11310	Porcine reproducti
37	30.6	9.2	648	21 A49675	HAL04-EST AA891494
38	30.6	9.2	649	21 A49674	HAL04-EST AA891494
39	30.6	9.2	758	21 C46428	Arabidopsis thalia
40	30.6	9.2	775	21 C35399	Arabidopsis thalia
41	30.6	9.2	1825	21 C44207	Human low adenosin
42	30.6	9.2	2163	21 F21411	Human adenosine re
43	30.6	9.2	2163	21 A35289	Human low adenosin
44	30.6	9.2	38644	21 F21424	Human adenosine re
45	30.6	9.2	38644	21 A35302	Human adenosine re

ALIGNMENTS

RESULT	1.
T42063	T42063 standard; DNA; 1830121 BP.
XX	
AC	T42063:
XX	
DT	14-SEP-1999 (first entry)
XX	
DE	Haemophilus influenzae complete genome sequence.
XX	
KW	Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW	expression modulating fragment; regulation; gene expression; vector;
KW	organism; open reading frame; ORF; ds.
XX	
OS	Haemophilus influenzae.
XX	
PN	W09633276-A1.
XX	
PD	24-OCT-1996.
XX	
PF	22-APR-1996; 96MO-US05320.
XX	
PR	07-JUN-1995; 95US-0487429.
PR	21-APR-1995; 95US-0426787.
PR	07-JUN-1995; 95US-0476102.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	(UYJO) UNIV JOHNS HOPKINS.
XX	
PI	Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX	WPI; 1996-485782/48.
XX	
PT	Haemophilus influenzae Rd genome recorded on computer readable

PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching

PS Claim 1; Page 77.2-77.1091; 1291pp; English.

XX This sequence represents the complete genome sequence of the bacterium
 CC Haemophilus influenzae strain Rd. The invention relates to a computer
 CC readable medium (CRM) having recorded upon it the complete H. influenzae
 CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
 CC sequence at least 99% identical to (1). By providing the full-length
 CC genomic sequence in a computer readable form, it is possible to identify
 CC commercially important nucleic acid fragments and expression modulating
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
 CC regulate the expression of a nucleic acid molecule. Vectors and altered
 CC organisms comprising the predicted ORFs can be used to produce any of the
 CC polypeptide fragments of the H. influenzae Rd genome.

SO Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other:

Query Match 11.5%; Score 38.2; DB 17; Length 1830121;
 Best Local Similarity 66.3%; Pred. No. 0.73;
 Matches 55; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 175 tggcaagaagagatgacccaacatgcaactttttattttttgtggtccatgtgc 234
 Db 1420637 tacgaagctggcagtaataaactgcacaattttcattcttctcctgttaa 1420696
 Qy 235 agagaagcacagacacatctgcc 257
 Db 1420697 agagtcgttaaaaaaacacatctgcc 1420719

RESULT 2
 X25344 X25344 standard; cDNA; 1443 BP.
 AC X25344;
 DT 19-JUL-1999 (first entry)
 XX Human adenine nucleotide translocator ANT5-related EST clone.
 DE
 XX ANT5; ANT; adenine nucleotide translocator; ADP/ATP translocator;
 KW human; therapy; diagnosis; congestive heart failure;
 KW ischaemic heart disease; arrhythmia; diastolic dysfunction;
 KW systolic dysfunction; hypertrophic cardiomyopathy; stroke; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 100..1443
 FT /*tag= a
 XX EP911341-A1.
 XX 28-APR-1999.
 PD 24-SEP-1998; 98EP-0203212.
 PF 02-JUL-1998; 98EP-0401655.
 PR 23-OCT-1997; 97EP-0402511.
 XX (SMIK) SMITHKLINE BEECHAM LAB PHARM.
 PA Brill AMA, Krief S, Souchet M;
 PI WPI; 1999-246386/21.
 DR P-PSDB; Y05713.
 XX New human calcium-sensitive adenine nucleotide translocator (ANT5)
 PT useful for diagnosing and treating congestive heart failure,
 PT ischemic heart disease and stroke

XX Claim 18; Page 18; 24pp; English.

PS This polynucleotide represents an isolated EST clone showing
 CC homology to a novel human calcium-sensitive adenine nucleotide
 CC translocator, termed ANT5 (see X25343). ANT5 polypeptides and
 CC polynucleotides (including the present sequence), recombinant
 CC materials, and methods for their production and use are claimed.
 CC ANT5 is thought to be instrumental in the regulation of energy
 CC fluxes by calcium, thus of muscle contractile function. ANT5
 CC polypeptides, polynucleotides, agonists and antagonists of the
 CC invention can be used to treat abnormal conditions related to ANT5
 CC activity or expression, such as congestive heart failure, ischemic
 CC heart disease, cardiac arrhythmia, diastolic or systolic
 CC dysfunction, hypertrophic cardiomyopathy and stroke. They can also
 CC be used in diagnostic assays to detect disease, or a susceptibility
 CC to disease, related to ANT5 expression or activity.

SQ Sequence 1443 BP; 356 A; 320 C; 371 G; 396 T; 0 other:

Query Match 11.1%; Score 37; DB 20; Length 1443;
 Best Local Similarity 54.9%; Pred. No. 0.083;
 Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 10 tgtccatccttggaaagaggtccacttgggtctcttaaggcgagatcccgctctc 69
 Db 827 tgtctgtcgtggcagccctgggttttttggatctacaagggtgcaaaagatgtcttc 886
 Qy 70 tctgattgtctctcttggaaagaaatgtgcaggtcatgagcttgcgaagaagca 129
 Db 887 tgcggagacatctctcttgcgcactacttccgtgtcatgtcgtgaagcttctc 946
 Qy 130 tgataaagatga 142
 Db 947 ttcgaatgaaaga 959

RESULT 3
 X25343 X25343 standard; cDNA; 2095 BP.
 AC X25343;
 DT 19-JUL-1999 (first entry)
 XX Human adenine nucleotide translocator ANT5 cDNA.
 DE
 XX ANT5; ANT; adenine nucleotide translocator; ADP/ATP translocator;
 KW human; therapy; diagnosis; congestive heart failure;
 KW ischaemic heart disease; arrhythmia; diastolic dysfunction;
 KW systolic dysfunction; hypertrophic cardiomyopathy; stroke; ss.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 71..2095
 FT /*tag= a
 XX EP911341-A1.
 XX 28-APR-1999.
 PD 24-SEP-1998; 98EP-0203212.
 PF 02-JUL-1998; 98EP-0401655.
 PR 23-OCT-1997; 97EP-0402511.
 XX (SMIK) SMITHKLINE BEECHAM LAB PHARM.
 PA Brill AMA, Krief S, Souchet M;
 PI WPI; 1999-246386/21.
 DR

DR P-PSDB; Y05712.
 XX
 PT New human calcium-sensitive adenine nucleotide translocator (ANT5)
 PT useful for diagnosing and treating congestive heart failure,
 PT ischemic heart disease and stroke
 XX
 PS Claim 9; Page 15-16; 24pp; English.
 CC This polynucleotide codes for a novel human calcium-sensitive
 CC adenine nucleotide translocator, termed ANT5 (see Y05712).
 CC Claimed polynucleotides comprising the present sequence, or
 CC homologous sequences, can be obtained from human heart, brain,
 CC uterus, mammary gland, lung, prostate, kidney, trachea, stomach,
 CC liver, placenta, testis, small intestine, spinal cord, ovary,
 CC spleen, pancreas, thymus, aorta, leukocyte, skeletal muscle,
 CC adrenal, adipose, lymph node, colon, thyroid, bone marrow, bladder,
 CC salivary gland or appendix cDNA. The invention relates to ANT5
 CC polypeptides and polynucleotides, recombinant materials and methods
 CC for their production and use. ANT5 may be instrumental in the
 CC regulation of energy fluxes by calcium, thus of muscle contractile
 CC function. The ANT5 polypeptides, polynucleotides, agonists and
 CC antagonists of the invention can be used to treat abnormal
 CC conditions related to ANT5 activity or expression, such as
 CC congestive heart failure, ischaemic heart disease, cardiac
 CC arrhythmia, diastolic or systolic dysfunction, hypertrophic
 CC cardiomyopathy and stroke. They can also be used in diagnostic
 CC assays to detect disease, or a susceptibility to disease,
 CC related to ANT5 expression or activity.
 XX
 SQ Sequence 2095 BP; 543 A; 460 C; 522 G; 570 T; 0 other;

Query Match 11.18; Score 37; DB 20; Length 2095;
 Best Local Similarity 54.9%; Pred. No. 0.098;
 Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 10 tttcacccttggaagaggtccacttggtccttaagggcgcatcccggtcttc 69
 DB 1479 tttcctgctggcgaggtcgtggttttttgggactcaagggcgcaagcatgcttc 1538
 QY 70 tctgattgctcctcttggaagcaatggaacttgagcatgagcttgcaagaagca 129
 DB 1539 tgcggagcatcctcttcgcgcacacttcctcgtgtatgcatgtaggaagcttc 1598
 QY 130 tgataaagaatga 142
 DB 1599 ttgcaaatgaaga 1611

RESULT 4
 C90467
 ID C90467 standard; cDNA; 2876 BP.
 AC C90467;
 XX
 XX
 DT 12-MAR-2001 (first entry)
 XX
 DE Human uncoupling protein cDNA #16.
 XX
 DE Human; uncoupling protein; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cardiatic; vasotropic;
 KW cerebroprotective; neuroprotective; antibacterial; ophthalmological;
 KW gastrointestinal; nephrotoxic; gynaecological; vulnary; thrombolytic;
 KW gene therapy; cancer; wound; infectious disease; thrombosis; arthritis;
 KW infertility; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200061614-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000MO-US09534.

XX 09-APR-1999; 99US-0128701.
 PR 08-JUL-1999; 99US-0142821.
 PR 18-AUG-1999; 99US-0149448.
 PR 12-NOV-1999; 99US-0164751.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Ni J, Komatsoulis G, Rosen CA, Soppet DR;
 XX
 DR WPI: 2000-656322/63.
 DR P-PSDB; B50393.
 XX
 PT Uncoupling proteins and nucleic acid sequences encoding them, useful
 PT for detecting, preventing and treating proliferative, neurological,
 PT immune system, cardiovascular and gastrointestinal disorders -
 XX
 PS Claim 1; Page 313-314; 343pp; English.
 CC The present sequence is one of eighteen isolated nucleotide sequences
 CC encoding uncoupling proteins. The nucleotide sequences may be used for
 CC the detection of various disorders such as cancer, for chromosome
 CC identification, as chromosome markers and for numerous other diagnostic
 CC or research purposes. The uncoupling protein encoded by the nucleotide
 CC sequences may be used to treat disorders such as neural, immune,
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
 CC renal and proliferative disorders, wounds, infectious diseases,
 CC thrombosis, arthritis, and infertility.
 XX
 SQ Sequence 2876 BP; 775 A; 695 C; 669 G; 737 T; 0 other;

Query Match 11.18; Score 37; DB 21; Length 2876;
 Best Local Similarity 54.9%; Pred. No. 0.11;
 Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 10 tttcacccttggaagaggtccacttggtccttaagggcgcatcccggtcttc 69
 DB 369 tttcctgctggcgaggtcgtggttttttgggactcaagggcgcaagcatgcttc 428
 QY 70 tctgattgctcctcttggaagcaatggaacttgagcatgagcttgcaagaagca 129
 DB 429 tgcggagcatcctcttcgcgcacacttcctcgtgtatgcatgtaggaagcttc 488
 QY 130 tgataaagaatga 142
 DB 489 ttgcaaatgaaga 501

RESULT 5
 C77456
 ID C77456 standard; cDNA; 3145 BP.
 AC C77456;
 XX
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3011 polynucleotide sequence SEQ ID NO:6021.
 XX
 DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiproliferative; antiparkinsonian; neurotoxic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiatic;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

```

XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX PF 31-MAR-2000; 2000WO-US08621.
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach M;
XX DR WPI: 2000-602362/57.
XX DR P-PSDB; B43247.
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 5; Page 5201-5203; 5507pp; English.
XX CC C74446 to C77606 encode the proteins given in B40237 to B43397, which
XX CC represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulneryary;
XX CC antiproliferic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;
XX CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;
XX CC cardiatic; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX CC dermatologic; immunosuppressive; antinflammatory; antibacterial;
XX CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX CC sequences can be used for determining the presence of or predisposition
XX CC to, or preventing or treating pathological conditions associated with an
XX CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX CC proteins in gene therapy vectors. The proteins and nucleic acids may be
XX CC used to treat cancers, proliferative disorders, neurodegenerative
XX CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX CC storage, systemic lupus erythematosus, severe combined immunodeficiency
XX CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
XX CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX SS Sequence 3145 BP; 874 A; 629 C; 716 G; 925 T; 1 other:
XX
Query Match 11.1%; Score 37; DB 21; Length 3145;
Best Local Similarity 54.9%; Pred. No. 0.12;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
OY 10 tgcacatccttggaagaggtccacttggtctttaaaggcgagatccccgctct 69
DB 1547 tgcctgcctgcgagacctggggcttcttggaactacaagggtccaaagatgctctc 1606
OY 70 tctgattgctcctcttggaagcaatcttgcaagctatgagcttgcaagagcgaa 129
DB 1607 tgcggagactctctcttcgcgcactactcttcgctgcatgctatggaaggtctcct 1666
OY 130 tgataaagatga 142
DB 1667 ttgcaaatgaaga 1679

```

```

RESULT 6
ID Z20025 standard; cDNA: 1267 BP.
XX Z20025;
AC Z20025;

```

```

XX DT 21-DEC-1999 (first entry)
XX DE Wheat brittle-1 partial cDNA.
XX KW Brittle-1; wheat; carbohydrate; starch; transgenic plant; ss.
XX OS Triticum aestivum.
XX FH Key Location/Qualifiers
XX FT 2..871
XX FT CDS /*tag= a
XX FT /partial
XX PN WO9949047-A2.
XX PD 30-SEP-1999.
XX PF 22-MAR-1999; 99WO-US06583.
XX PR 26-MAR-1998; 98US-0079420.
XX PA (DUPO ) DU POWT DE NEMOURS & CO E. I.
XX PI Allen SM, Hitz WD, Lightner JE, Rafalski JA;
XX DR WPI: 1999-591098/50.
XX DR P-PSDB; Y31936.
XX PT Novel genes useful in studies of carbohydrate metabolism and function
XX PT in plants -
XX PS Claim 7; Page 42; 42pp; English.
XX CC This nucleotide sequence represents a portion of the cDNA insert in
XX CC clone wreln.pk0049.e1 encoding a portion (see Y31936) of wheat
XX CC brittle-1, a plastidic membrane transporter involved in the
XX CC transport of ADP-glucose from the cytosol to the plastid where it
XX CC is used for starch biosynthesis. The clone was isolated from a
XX CC wheat root cDNA library. The invention relates to isolated
XX CC nucleic acid fragments (see Z20012-25) encoding plant carbohydrate
XX CC biosynthetic enzymes (see Y31923-36) selected from 1,3-beta-D-glucan
XX CC synthase and brittle-1. It also relates to the construction of a
XX CC chimeric gene encoding all or a portion of a carbohydrate
XX CC biosynthetic enzyme, in sense or antisense orientation, where
XX CC expression of the chimeric gene results in altered levels of
XX CC carbohydrate biosynthetic enzyme in a transformed host cell. The
XX CC availability of nucleic acids encoding these enzymes will facilitate
XX CC studies of carbohydrate metabolism and function in plants, provide a
XX CC genetic tools for the manipulation of these pathways, and provide a
XX CC means to control starch and 1,3-beta-D-glucan biosynthesis in plant
XX CC cells.
XX SS Sequence 1267 BP; 351 A; 268 C; 317 G; 331 T; 0 other:
XX
Query Match 10.9%; Score 36.4; DB 20; Length 1267;
Best Local Similarity 52.7%; Pred. No. 0.12;
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
OY 7 tctctcactccttggaagcgaggtccacttggtctttaaaggcgagatccccgct 66
DB 726 tcttgaccatcttcgagcgagggttggtgggtctcaccagagactggggcttgct 785
OY 67 tctctcgtatgctcctcttgagcaatgaacttgcaagctatgagcttgcaagaag 126
DB 786 gatgaagctggtgctcgtctgcttggtgattcgcttatgctgacgaagcttgcaagaaga 845
OY 127 caatgataaagatgagaanaattccagtg 156
DB 846 tactgattgaggaagaagaacgaatgaagcg 875

```


DT	19-MAR-1999	(first entry)
XX		
DE	Enterococcus faecalis genome contig SEQ ID NO:100.	
XX		
KM	Enterococcus faecalis; contig; detection; Enterococcal infection;	
KW	vaccine; attenuation; computer readable medium; ds.	
XX		
OS	Enterococcus faecalis.	
PN		
XX	WO9850555-A2.	
PD		
XX	12-NOV-1998.	
PF		
XX	04-MAY-1998; 98WO-US08965.	
PR		
XX	14-NOV-1997; 97US-0066009.	
PR	06-MAY-1997; 97US-0044031.	
PR	16-MAY-1997; 97US-0046655.	
PA		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PI		
XX	Barash SC, Dillon PJ, Kunsch CA;	
DR		
XX	WPI; 1999-045171/04.	
PT		
XX	New isolated Enterococcus faecalis polynucleotides and polypeptides	
PT	- used to develop products for the detection of Enterococcus and for	
PT	use in vaccines for prevention or attenuation of Enterococcus	
XX	infection.	
PS		
XX	Claim 1; Page 645-661; 2084pp; English.	
XX		
CC	A computer readable medium has been developed which has recorded on it	
CC	982 nucleotide sequences isolated from the Enterococcus faecalis genome.	
CC	X12938 to X13919 represent these nucleotide sequences which are primary	
CC	nucleotide sequences, also known as contigs. The computer-based system	
CC	can identify fragments of the Enterococcus faecalis genome with	
CC	commercial importance. The products can be used to detect the presence	
CC	of Enterococcus faecalis in samples. They can also be used for	
CC	diagnosing Enterococcal infection in an animal and monitoring	
CC	progression of disease, and for identifying agents which can be used to	
CC	modulate the growth or pathogenicity of Enterococcus faecalis, or	
CC	another related organism, in vivo or in vitro. In particular the	
CC	polypeptides encoded by the Enterococcus faecalis nucleotide sequences	
CC	can be used in vaccines to prevent or attenuate an Enterococcal	
XX	infection.	
SQ		
Sequence	32768 BP; 9063 A; 7056 C; 5292 G; 11350 T; 7 other:	
Query Match	10.2%; Score 33.8; DB 20; Length 32768;	
Best Local Similarity	52.5%; Pred. No. 3.1;	
Matches 74; Conservative	0; Mismatches 67; Indels 0; Gaps 0;	
OY	78 gctcctcttggagcaatggaactttgcaggcgataagcttcgccagaaggaacgtataaaa 137 	
Db	30997 gctagtitttgagcatcttaagtgctgcgcctatgatgcgcataatgaataatgaacgyc 31056	
OY	138 gatgagaaaaattccagtgagtcacaacgcagagagatcttcgccagaagacgtagtacacc 197 	
Db	31057 aacctcaaaatttggccaattggaatagagcaaacattagaagaagaanaatttatcacgc 31116	
OY	198 atgcaactttttttttttttttt 218 	
Db	31117 aaaccttttaatgaagcaatt 31137	
RESULT	9	
ID	F09810	
XX	F09810 standard; cDNA; 1172 BP.	
XX		
AC	F09810;	
XX		

PR	17-JUN-1999;	9905-0133492;
PR	18-JUN-1999;	9905-0133454;
PR	18-JUN-1999;	9905-0133455;
PR	18-JUN-1999;	9905-0133456;
PR	18-JUN-1999;	9905-0133457;
PR	18-JUN-1999;	9905-0133458;
PR	18-JUN-1999;	9905-0133459;
PR	18-JUN-1999;	9905-0133460;
PR	18-JUN-1999;	9905-0133461;
PR	18-JUN-1999;	9905-0133462;
PR	21-JUN-1999;	9905-0133817;
PR	22-JUN-1999;	9905-0133819;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140695;
PR	28-JUN-1999;	9905-0140923;
PR	29-JUN-1999;	9905-0140981;
PR	30-JUN-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141842;
PR	01-JUL-1999;	9905-0141854;
PR	02-JUL-1999;	9905-0142055;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0142920;
PR	12-JUL-1999;	9905-0142977;
PR	13-JUL-1999;	9905-0143542;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144005;
PR	15-JUL-1999;	9905-0144085;
PR	16-JUL-1999;	9905-0144325;
PR	19-JUL-1999;	9905-0144331;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144333;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144352;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144684;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0144816;
PR	21-JUL-1999;	9905-0144808;
PR	21-JUL-1999;	9905-0145085;
PR	22-JUL-1999;	9905-0145085;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145192;
PR	22-JUL-1999;	9905-0145198;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145224;
PR	26-JUL-1999;	9905-01453276;
PR	27-JUL-1999;	9905-0145913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146388;
PR	03-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147202;
PR	04-AUG-1999;	9905-0147304;
PR	05-AUG-1999;	9905-0147192;
PR	05-AUG-1999;	9905-0147260;
PR	06-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147935;
PR	11-AUG-1999;	9905-0148171;
PR	11-AUG-1999;	9905-0148171;
PR	13-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148565;

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.9%; Score 33; DB 21; Length 1195;
Best Local Similarity 55.8%; Pred. No. 1.3;

Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Oy 125 ggcgaatgataaagatgtggaataattccagtgcgaacagagagatccttgcgaagaa 184
Db 1007 ggcgattacaaactggtctctgtcacatctatagtgtgagggagatgttcaagaag 1066
Oy 185 gcagtgatcaacatgcgaactttttatttattttttgtggtcagatgtgcaga 237

DB 1067 attcgaatcaatctcgggattttcagtttcttactgtctatcttccana 1119
RESULT 12
C33296
ID C33296 standard; DNA; 1198 BP.
XX C33296;
AC C33296;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2508.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
PM
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136787.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

```
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139859.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 20-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144684.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.

PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151085.
PR 27-AUG-1999: 99US-0151086.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 22-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 25-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 26-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 9.9%; Score 33; DB 21; Length 1198;
Best Local Similarity 55.8%; Pred. NO. 1.3;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```


PR 05-MAR-1999;	9905-0123180.
PR 09-MAR-1999;	9905-0123548.
PR 23-MAR-1999;	9905-0125788.
PR 25-MAR-1999;	9905-0126264.
PR 29-MAR-1999;	9905-0126785.
PR 01-APR-1999;	9905-0127462.
PR 06-APR-1999;	9905-0128234.
PR 16-APR-1999;	9905-0128714.
PR 19-APR-1999;	9905-0129845.
PR 21-APR-1999;	9905-0130077.
PR 23-APR-1999;	9905-0130449.
PR 28-APR-1999;	9905-0130510.
PR 30-APR-1999;	9905-0130891.
PR 30-APR-1999;	9905-0131449.
PR 30-APR-1999;	9905-0132048.
PR 04-MAY-1999;	9905-0132407.
PR 05-MAY-1999;	9905-0132484.
PR 06-MAY-1999;	9905-0132486.
PR 07-MAY-1999;	9905-0132487.
PR 11-MAY-1999;	9905-0132863.
PR 14-MAY-1999;	9905-0134256.
PR 14-MAY-1999;	9905-0134218.
PR 14-MAY-1999;	9905-0134219.
PR 14-MAY-1999;	9905-0134221.
PR 18-MAY-1999;	9905-0134370.
PR 19-MAY-1999;	9905-0134768.
PR 20-MAY-1999;	9905-0134941.
PR 21-MAY-1999;	9905-0135114.
PR 24-MAY-1999;	9905-0135353.
PR 25-MAY-1999;	9905-0135629.
PR 27-MAY-1999;	9905-0136021.
PR 28-MAY-1999;	9905-0136392.
PR 01-JUN-1999;	9905-0136782.
PR 03-JUN-1999;	9905-0137528.
PR 04-JUN-1999;	9905-0137502.
PR 07-JUN-1999;	9905-0137724.
PR 08-JUN-1999;	9905-0138094.
PR 10-JUN-1999;	9905-0138540.
PR 14-JUN-1999;	9905-0138847.
PR 16-JUN-1999;	9905-0139452.
PR 17-JUN-1999;	9905-0139453.
PR 18-JUN-1999;	9905-0139482.
PR 18-JUN-1999;	9905-0139454.
PR 18-JUN-1999;	9905-0139455.
PR 18-JUN-1999;	9905-0139456.
PR 18-JUN-1999;	9905-0139457.
PR 18-JUN-1999;	9905-0139458.
PR 18-JUN-1999;	9905-0139459.
PR 18-JUN-1999;	9905-0139460.
PR 18-JUN-1999;	9905-0139461.
PR 18-JUN-1999;	9905-0139462.
PR 18-JUN-1999;	9905-0139463.
PR 18-JUN-1999;	9905-0139750.
PR 18-JUN-1999;	9905-0139763.
PR 21-JUN-1999;	9905-0139817.
PR 22-JUN-1999;	9905-0139899.
PR 23-JUN-1999;	9905-0140353.
PR 23-JUN-1999;	9905-0140354.
PR 24-JUN-1999;	9905-0140695.
PR 28-JUN-1999;	9905-0140823.
PR 29-JUN-1999;	9905-0140921.
PR 30-JUN-1999;	9905-0141287.
PR 01-JUL-1999;	9905-0141842.
PR 01-JUL-1999;	9905-0142154.
PR 02-JUL-1999;	9905-0142055.
PR 06-JUL-1999;	9905-0142350.
PR 08-JUL-1999;	9905-0142803.
PR 09-JUL-1999;	9905-0142920.
PR 12-JUL-1999;	9905-0142977.
PR 13-JUL-1999;	9905-0143542.
PR 14-JUL-1999;	9905-0143624.
PR 15-JUL-1999;	9905-0144005.
PR 16-JUL-1999;	9905-0144085.
PR 16-JUL-1999;	9905-0144086.
PR 19-JUL-1999;	9905-0144325.
PR 19-JUL-1999;	9905-0144331.
PR 19-JUL-1999;	9905-0144332.
PR 19-JUL-1999;	9905-0144333.
PR 19-JUL-1999;	9905-0144334.
PR 19-JUL-1999;	9905-0144335.
PR 20-JUL-1999;	9905-0144352.
PR 20-JUL-1999;	9905-0144632.
PR 21-JUL-1999;	9905-0144884.
PR 21-JUL-1999;	9905-0144884.
PR 21-JUL-1999;	9905-0145086.
PR 21-JUL-1999;	9905-0145088.
PR 22-JUL-1999;	9905-0145085.
PR 22-JUL-1999;	9905-0145087.
PR 22-JUL-1999;	9905-0145089.
PR 22-JUL-1999;	9905-0145192.
PR 23-JUL-1999;	9905-0145145.
PR 23-JUL-1999;	9905-0145218.
PR 26-JUL-1999;	9905-0145224.
PR 27-JUL-1999;	9905-0145276.
PR 27-JUL-1999;	9905-0145913.
PR 27-JUL-1999;	9905-0145918.
PR 28-JUL-1999;	9905-0145919.
PR 02-AUG-1999;	9905-0146386.
PR 02-AUG-1999;	9905-0146388.
PR 02-AUG-1999;	9905-0146389.
PR 03-AUG-1999;	9905-0147038.
PR 04-AUG-1999;	9905-0147204.
PR 05-AUG-1999;	9905-0147302.
PR 05-AUG-1999;	9905-0147312.
PR 06-AUG-1999;	9905-0147260.
PR 06-AUG-1999;	9905-0147303.
PR 06-AUG-1999;	9905-0147416.
PR 09-AUG-1999;	9905-0147493.
PR 09-AUG-1999;	9905-0147935.
PR 10-AUG-1999;	9905-0148171.
PR 11-AUG-1999;	9905-0148319.
PR 12-AUG-1999;	9905-0148341.
PR 13-AUG-1999;	9905-0148565.
PR 13-AUG-1999;	9905-0148684.
PR 16-AUG-1999;	9905-0149368.
PR 17-AUG-1999;	9905-0149175.
PR 18-AUG-1999;	9905-0149426.
PR 20-AUG-1999;	9905-0149722.
PR 20-AUG-1999;	9905-0149723.
PR 20-AUG-1999;	9905-0149929.
PR 23-AUG-1999;	9905-0149902.
PR 23-AUG-1999;	9905-0149930.
PR 25-AUG-1999;	9905-0150566.
PR 26-AUG-1999;	9905-0150884.
PR 27-AUG-1999;	9905-0151085.
PR 27-AUG-1999;	9905-0151066.
PR 27-AUG-1999;	9905-0151080.
PR 30-AUG-1999;	9905-0151303.
PR 31-AUG-1999;	9905-0151438.
PR 01-SEP-1999;	9905-0151930.
PR 07-SEP-1999;	9905-0152363.
PR 10-SEP-1999;	9905-0153070.
PR 13-SEP-1999;	9905-0153758.
PR 15-SEP-1999;	9905-0154018.
PR 16-SEP-1999;	9905-0154039.
PR 20-SEP-1999;	9905-0154779.
PR 22-SEP-1999;	9905-0155119.
PR 23-SEP-1999;	9905-0155486.
PR 24-SEP-1999;	9905-0155659.
PR 28-SEP-1999;	9905-0156458.
PR 29-SEP-1999;	9905-0156596.
PR 04-OCT-1999;	9905-0157117.

```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match

Best Local Similarity 9.7%; Score 32.2; DB 21; Length 1289;

Matches 61; Conservative 56.0%; Pred. No. 2.5; Mismatches 48; Indels 0; Gaps 0;

```
Qy 22 gaacgaggggtccacttggtcttcaaggcgcgatccccggctctcttgattgttc 81
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 493 gaaccgagggactcgtgactgttcaaggaaatgacccaactgtgtcgtatgttc 552
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 82 ctcttggaagcaatgaacttgaggtatgagcttgccaagaaggaat 130
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 553 ctaatcggtgtttaattttcagctatgagcaagcttcaaatgcat 601
```

Search completed: May 8, 2001, 15:17:47
Job time: 43608 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:19 ; Search time 132.33 Seconds
(Without alignments)
439.377 Million cell updates/sec

Title: US-09-619-643-5

Perfect score: 333

Sequence: 1 tgaatactgtcaccctt.....acattacatgagaagaatgc 333

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	34.8	10.5	10627	1 US-08-060-925A-12	Sequence 12, Appl
2	31.8	9.5	1816	4 US-09-188-930-262	Sequence 262, App
3	29.4	8.8	3505	1 US-07-660-465-1	Sequence 1, Appl
4	29.2	8.8	1949	2 US-08-937-466-3	Sequence 3, Appl
5	29.2	8.8	1949	2 US-08-172-528-3	Sequence 3, Appl
6	29.2	8.8	1949	3 US-09-318-199-3	Sequence 3, Appl
C 7	29	8.7	3765	4 US-07-705-490-1	Sequence 1, Appl
C 8	29	8.7	3765	4 US-07-751-891B-1	Sequence 1, Appl
C 9	29	8.7	4362	2 US-08-455-073A-1	Sequence 1, Appl
C 10	28.8	8.6	5874	4 US-08-843-417-9	Sequence 9, Appl
C 11	28.6	8.6	1258	3 US-09-155-200-3	Sequence 3, Appl
12	28.4	8.5	1021	1 US-08-198-446B-2	Sequence 2, Appl
13	28.4	8.5	1021	2 US-08-870-693-2	Sequence 2, Appl
14	28.2	8.5	3073	1 US-07-688-352C-31	Sequence 31, Appl
15	28.2	8.5	3073	2 US-08-474-379C-31	Sequence 31, Appl
16	28.2	8.5	3073	3 US-09-146-249A-31	Sequence 31, Appl
17	28.2	8.5	3073	4 US-08-206-188B-31	Sequence 31, Appl
18	28.2	8.5	3073	5 PCT-US91-02714-30	Sequence 30, Appl
19	28	8.4	1592	1 US-08-044-618-4	Sequence 4, Appl
20	28	8.4	2682	1 US-08-044-618-3	Sequence 3, Appl
21	28	8.4	3747	1 US-08-044-618-5	Sequence 5, Appl
22	28	8.4	5345	1 US-08-044-618-7	Sequence 7, Appl
C 23	28	8.4	7400	1 US-07-674-852-1	Sequence 1, Appl
C 24	28	8.4	7400	4 US-08-473-185-1	Sequence 1, Appl
C 25	27.8	8.3	518	1 US-08-485-284A-2	Sequence 2, Appl
C 26	27.8	8.3	807	2 US-08-531-927B-9	Sequence 9, Appl
C 27	27.8	8.3	1776	2 US-08-531-927B-1	Sequence 1, Appl

28	27.8	8.3	2782	2 US-08-937-466-1	Sequence 1, Appl
29	27.8	8.3	2782	2 US-09-172-528-1	Sequence 1, Appl
30	27.8	8.3	2782	3 US-09-318-199-1	Sequence 1, Appl
31	27.8	8.3	4507	2 US-08-568-459A-3	Sequence 3, Appl
32	27.8	8.3	4507	2 US-08-467-826B-3	Sequence 3, Appl
C 33	27.6	8.3	907	2 US-08-486-663A-12	Sequence 12, Appl
C 34	27.6	8.3	907	3 US-08-767-942A-12	Sequence 12, Appl
C 35	27.6	8.3	907	5 PCT-US95-06722-23	Sequence 23, Appl
C 36	27.6	8.3	1192	4 US-09-142-565-1	Sequence 1, Appl
C 37	27.6	8.3	2715	2 US-08-359-705B-5	Sequence 5, Appl
C 38	27.6	8.3	2715	2 US-08-286-846A-5	Sequence 5, Appl
C 39	27.6	8.3	2715	2 US-08-457-880A-5	Sequence 5, Appl
C 40	27.6	8.3	2715	3 US-08-444-622A-5	Sequence 5, Appl
C 41	27.6	8.3	2715	3 US-08-942-562-5	Sequence 5, Appl
C 42	27.6	8.3	2715	4 US-09-156-923-5	Sequence 5, Appl
C 43	27.6	8.3	3501	1 US-08-524-757-5	Sequence 5, Appl
C 44	27.6	8.3	15101	2 US-08-799-464A-14	Sequence 14, Appl
C 45	27.6	8.3	15101	5 PCT-US95-09927-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-060-925A-12/C
Sequence 12, Application US/08060925A
Patent No. 5439824
GENERAL INFORMATION:
APPLICANT: Brantley, Mark
APPLICANT: Laubach, Victor
TITLE OF INVENTION: INCREASED EXPRESSION OF ALPHA-1
TITLE OF INVENTION: ANTITRYPSIN IN EXPRESSION VECTORS THROUGH THE INCLUSION OF
NUMBER OF INVENTION: INTRON II
NUMBER OF SEQUENCE ADDRESSES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: KNOBE, MARTENS, OLSON AND BEAR
STREET: 620 NEWPORT CENTER DRIVE SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/060, 925A
FILING DATE: 06-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael L.
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH040.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 10627 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-060-925A-12

Query Match 10.5%; Score 34.8; DB 1; Length 10627;
Best Local Similarity 50.0%; Pred. No. 0.27;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY	72	tgagatgctctctcttgagcaagaacatttgagagctatgagcttgcagaagaagcaatg	131
Db	8870	TGGCTCGGGCTCTCGATCCGTACCCCTCTAGCCCTTTGTCATTCATCAACCCGTGG	8811
QY	132	ataaagaatgagaaaattccagctgagctcaacagcagagactcttcgcaagaagaagcaga	191
Db	8810	AGCAGAGACCTCTAGAGAGTCTGTCATGTGACACAGAAATTCACGCTCGATTCCTATTA	8751
QY	192	tcaaccatgcaactttttattttatttatttcttttggtgccatgctgcagagaagaac	245
Db	8750	TGAACCCGAGACCTTTTGATTTTCACTCTGGTTTATACGTGTACAAACGCAAC	8697

```

RESULT 2
US-09-188-930-262
: Sequence 262, Application US/09188930A
: Patent No. 6150502
:
: GENERAL INFORMATION:
:
: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorina
: APPLICANT: Sleeman, Mathew
: APPLICANT: Onrust, Rene
: APPLICANT: Murlson, James Greg
:
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
:
: TITLE OF INVENTION: and Methods For Their Use
:
: FILE REFERENCE: 11000.1011c1
:
: CURRENT APPLICATION NUMBER: US/09/188,930A
:
: CURRENT FILING DATE: 1998-11-09
:
: NUMBER OF SEQ. ID NOS: 148
:
: SOFTWARE: FastSeq for Windows Version 3.0
:
: SEQ. ID NO 262
:
: LENGTH: 1816
:
: TYPE: DNA
:
: ORGANISM: Mouse
:
: US-09-188-930-262

```

Query Match	9.5%	Score 31.8	DB 4	Length 1816
Best Local Similarity	57.6%	Pred. No. 1.1		
Matches 57, Conservative	0	Mismatches	42	Indels 0
				Gaps 0

QY	15	atcccttgaaacagagggtccaaacttgggcctttaaaaggcgatcccggtcttctcgg	74
Db	1037	atcttggcctaaagagggtgttagctgctctctaaagaagctaaatcccacatgtctggg	1036
QY	75	atgtgctcctcttggagcaatgaacttgcagctatgag	113
Db	1097	atcatccctatgtctgcagcatcgcccaagctgtctatag	1135

RESULT 3
 US-07-660-465-1
 ; Sequence 1, Application US/07660465
 ; Patent No. 5176997
 ; GENERAL INFORMATION:
 ; APPLICANT: Humphreys-Beher, Michael G.
 ; TITLE OF INVENTION: DNA Probe for Male Infertility
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Saliwanchik & Saliwanchik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: FL
 ; COUNTRY: USA
 ; ZIP: 32606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/660,465
 ; FILING DATE: 19910222

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/371,472
FILING DATE: 26-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: Salliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: Uf/S&S-104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3505 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-660-463-1

```

Query Match	8.88;	Score 29.4;	DB 1;	Length 3505;
Best Local Similarity	47.58;	Pred. NO. 8;		
Matches 87; Conservative	0;	Mismatches 96;	Indels 0;	Gaps 0;

[illegible]

Qy	201	caa	203
		!!!	
Db	344	CAA	346

RESULT 4
 US-08-937-466-3
 : Sequence 3, Application US/08937466
 : Patent No. 5846779
 : GENERAL INFORMATION:
 : APPLICANT: Zhang, Ning
 : APPLICANT: Amaral, M. Catherine
 : APPLICANT: Chen, Jin-Long
 : TITLE OF INVENTION: UCP3 Genes
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
 : STREET: 75 DENISE DRIVE
 : CITY: HILLSBOROUGH
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 94010
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/937,466
 : FILING DATE:
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: OSMAN, RICHARD A
 : REGISTRATION NUMBER: 36,627
 : REFERENCE/DOCKET NUMBER: T97-009
 : TELECOMMUNICATION INFORMATION:

Oy	66	ttctctcggatctgcctcctcttgagcaatgaacttgcagcgagcatbagcttgcgaagaag	125
Db	1065	TTTCTGGCTGGGAGCTTGGACGATGATGTTTGTAACTATGACGCAACTGAAGAGC	1122
Oy	126	gcaatgataaagaatgag	143
Db	1125	GCCTTAATGAAAAGTCGAG	1142

```

RESULT 7
US-07-705-490-1/c
; Sequence 1, Application US/07705490
; Patent No. 6107025
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Nelson, David L.
; APPLICANT: Pieretti, Maura
; APPLICANT: Warren, Stephen T.
; APPLICANT: Oostla, Ben A.
; TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas D. Paul
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/705,490
; FILING DATE: 19910708
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3765 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-07-705-490-1

```

[illegible]

Patent No. 6180337
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.,
Nelson, David I.,
Pieretti, Maura
Warren, Stephen T.,
Oostra, Ben A.
Fu, Ying-hui
TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas D. Paul
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/751,891B
FILING DATE: 29-Aug-1991
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5350
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3765 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-07-751-891B-1

```

Query Match      8.7%; Score 29; DB 4; Length 3765;
Best Local Similarity 54.1%; Pred. No. 11;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0

QY 172 tcttgccaagaagcagtgatcacacatgcacattttatttatttcttgggtccatg 231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2591 TGTATCCCAATTAATTCCTCCCTGACACTGCTCATCCCAATTAATTAATTCCTACCTGCCAAGA 2532
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 232 tgcgagaaagcaagacacgtcgccttttttgatgatgggaattaca 280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2531 TATATAAAGCACAGATCATACAACTTTTGTGTAAGAAATGACA 2483
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-455-073A-1/C
: Sequence 1, Application US/08455073A
: Patent No. 5876949
: GENERAL INFORMATION:
: APPLICANT: Gideon Dreyfuss
: APPLICANT: Mikiko C. Sloml
: APPLICANT: Yan Zhang
: TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods
: TITLE OF INVENTION: Of Making And Using The Same
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949-15

```


Db 849 CGACAAACCATCTTCGCTTCTGTCGGGTTCTCAATGAAGACTCAGACAGTTTTCACC 790

Db 147 aattccagtgagtcacacgcgagatcttcgcagaagaacagc 189
||| ||| | ||| ||||| |||||
789 CATTGCAGACTCAACAGCTGATTTCACTTCGAAGAACCAGT 747

```

US-08-198-446B-2
US-08-198-446B-2
Sequence 2, Application US/08198446B
Patent No. 5674996
GENERAL INFORMATION:
APPLICANT: Hartwell, Ireland H.
APPLICANT: Weinert, Ted A.
APPLICANT: Plon, Sharon E.
APPLICANT: Groudine, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/198,446B
FILING DATE: 18-FEB-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI7537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: Yeast RAD24 cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Saccharomyces cerevisiae
US-08-198-446B-2

```

Query Match	8.5%	Score 28.4	DB 1	Length 1021	
Best Local Similarity	56.4%	Pred. No. 9.8	41	Indels	0
Matches 53	Conservative 0	Mismatches	0	Gaps	0
QY	205	tttttaatttatttttgggtccagtggttcagagaagcacagaacatgcctttttt	264		
Db	295	TTATTAAATGTTATTTCACAGGCGCCGTGCCAATATCCTTAGCACAATATTACACATCATG	354		
QY	265	gatatgggaattacatccgctagggcctctag	298		
Db	355	GATTTTGCTTGATGATATTCCGACATCCCTTTGAG	388		
RESULT 13					
US-08-870-693-2					
; Sequence 2, Application US/08870693					
Patent No. 5866338					

```

GENERAL INFORMATION:
APPLICANT: Hartwell, Leland H.
APPLICANT: Weinert, Ted A.
APPLICANT: Plon, Sharon E.
APPLICANT: Grounne, Mark T.
TITLE OF INVENTION: Cell Cycle Checkpoint Genes
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,693
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,446
FILING DATE: February 18, 1994
APPLICATION NUMBER: PCT/US93/04458
FILING DATE: May 12, 1993
APPLICATION NUMBER: US 07/884,426
FILING DATE: May 14, 1992
APPLICATION NUMBER: US 07/882,051
FILING DATE: May 12, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Shelness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCRI10798
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-224-0779
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: yeast RAD24 cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
US-08-870-693-2

```

```

Query Match      8.5%; Score 28.4; DB 2; Length 1021;
Best Local Similarity 56.4%; Pred. No. 9.8;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 205 ttcttatttatttttggtgcacatggtcagagaagacagaacatcgtcttttt 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 TTATTAAAGTTATTCACAGAGCCCTGTCCCATATCTTAGACATATTCTACCACTCAG 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 265 gatattggaattacatctccgttagggccttgag 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 GATTTTGGTTGATTAATTTCCGCAATCCCTTGTGAG 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-07-688-352C-31
; Sequence 31, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wiegler, Michael H.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:35 ; Search time 3095.41 Seconds
(without alignments)
939.818 Million cell updates/sec

Title: US-09-619-643-5

Perfect score: 333

Sequence: 1 tgataactctgtccatcctt.....acattacatgagaaagtgc 333

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

```
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*
```

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	331.4	99.5	442	104	AI978199
2	329.8	99.0	582	24	AI734448
3	288.8	86.7	385	110	AM067405
4	139	41.7	552	110	AM067175
5	137.4	41.3	499	144	BF177094
6	134.2	40.3	528	151	BE656155
7	71.6	21.5	319	30	AV523385
8	71.6	21.5	443	30	AV523684
9	71.6	21.5	572	30	AV561206
10	71.6	21.5	628	30	AV441249
11	67.6	20.3	468	146	BF278346
12	67.6	20.3	468	146	BF278348
13	65.6	19.7	790	137	BE642959
14	64.6	19.4	367	141	BE919380
15	57.2	17.2	315	104	AI998007
16	52.8	15.9	487	137	BE611103
17	50.2	15.1	458	107	AU166657
18	40.2	12.1	700	31	AV702549

C	19	39.6	11.9	217	23	A1670671	A1670671 605037F10
C	20	39.4	11.8	598	23	A1691263	A1691263 606028D02
	21	37.8	11.4	297	16	A1154643	A1154643 u037a06.r
	22	37.8	11.4	451	9	AA560947	AA560947 v170f04.r
	23	37.8	11.4	482	16	A1155074	A1155074 u078f08.r
	24	37.8	11.4	613	166	BE373592	BE373592 601225976
	25	37.8	11.4	695	141	BE912477	BE912477 601666103
	26	37.8	11.4	710	144	BE161364	BE161364 601770114
	27	37.8	11.4	729	147	BF385055	BF385055 602045653
	28	37.8	11.4	1008	145	BF235639	BF235639 602025569
	29	37.4	11.2	347	20	A1443882	A1443882 s444f04.y
	30	37.4	11.2	368	148	BF424070	BF424070 s14b12.y
	31	37.4	11.2	455	158	W76821	W76821 me73f10.r1
	32	37.4	11.2	473	105	AL386661	AL386661 MCB36803
	33	37.4	11.2	480	137	BE624311	BE624311 u043a06.y
	34	37.4	11.2	670	16	A1119522	A1119522 u104h04.y
C	35	37.2	11.2	446	173	BC079054	BC079054 H303C06-
C	36	37.2	11.2	628	208	AO578474	AO578474 nbx0093A
	37	37.2	11.1	843	172	BG029576	BG029576 602296749
	38	36.6	11.0	508	144	BF156894	BF156894 f158c04.y
	39	36.2	10.9	375	175	D22033	D22033 R1CC10212A
	40	36	10.8	402	202	AQ136330	AQ136330 HS_3056-B
	41	36	10.8	428	162	BE049964	BE049964 za54d01.b
	42	36	10.8	561	137	BE599547	BE599547 P11.88.G1
C	43	35.8	10.8	926	229	CNS016D0	AV379220 AV379220
	44	35.6	10.7	220	29	AV379220	AV379220 AV379220
	45	35.6	10.7	471	20	A1404883	A1404883 GH24733.5

ALIGNMENTS

RESULT 1
LOCUS A1978199/c
DEFINITION 614041D10.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
mRNA sequence.
ACCESSION A1978199
VERSION A1978199.1 GI:5791407
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
; Andropogoneae; Zea.
AUTHORS Walbot,V.
TITLE 1 (bases 1 to 442)
JOURNAL Zea mays ESTs from various cDNA libraries sequenced at Stanford
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614041 row: D column: 10.
Location/Qualifiers
source
1. 442
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="XLOLR"
/note="Organ: root; Vector: pluescriptII SK+; Site_1:
; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)"

BASE COUNT
ORIGIN

126 a 107 c 93 g 116 t

Query Match 99.5%; Score 331.4; DB 104; Length 442;
Best Local Similarity 99.7%; Pred. No. 5e-85;
Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy	1	tgaatacttccatcccttggaacagaggtccacttggcctctttaaaggcgcgctcc	60
Db	400	tgatratatttccatcccttggaacagaggtccacttggcctctttaaaggcgcgctcc	341
Oy	61	ccgctcttcttgatctgcctctcttggaacatgaatttgcaggctatagcttcca	120
Db	340	ccgctcttcttgatctgcctctcttggaacatgaatttgcaggctatagcttcca	281
Oy	121	agaagagcaatgaataagaatgaataatccagtgagtcacaacagagatcttgc	180
Db	280	agaagagcaatgaataagaatgaataatccagtgagtcacaacagagatcttgc	221
Oy	181	gaagcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga	240
Db	220	gaagcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga	161
Oy	241	gaacagacatctgcctcttcttgaatgaatgaatgaatgaatgaatgaatgaat	300
Db	160	gaacagacatctgcctcttcttgaatgaatgaatgaatgaatgaatgaatgaat	101
Oy	301	gactgtatgatatcatcatcatgaagaagtgc	333
Db	100	gactgtatgatatcatcatcatgaagaagtgc	68

RESULT 2
LOCUS A1734448/c
DEFINITION 606031E07.x1 606 - Ear tissue cDNA library from Schmidt Lab Zea
mays cDNA, mRNA sequence.
ACCESSION A1734448
VERSION A1734448.1 GI:5055561
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Mangoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
; Andropogoneae; Zea.
AUTHORS Walbot,V.
TITLE 1 (bases 1 to 582)
JOURNAL Zea mays ESTs from various cDNA libraries sequenced at Stanford
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606031 row: E column: 07.
Location/Qualifiers
source
1. 582
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
Lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
Lab"

BASE COUNT
ORIGIN

158 a 131 c 122 g 170 t 1 others

Query Match 99.0%; Score 329.8; DB 24; Length 582;
Best Local Similarity 99.4%; Pred. No. 1.5e-84;
Matches 331; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tgaatactctgcatcctcttgaaagaggtgacacttgagcctttaaaggcgatcc 60
|||||
DB 487 tcatatcttctgcatcctcttgaaagaggtgacacttgagcctttaaaggcgatcc 428
|||||
QY 61 ccggtctctcttgatctgctcctcttgagcaatgacacttgcaaggctatgaagcttgca 120
|||||
DB 427 ccggtctctcttgatctgctcctcttgagcaatgacacttgcaaggctatgaagcttgca 368
|||||
QY 121 aagaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 180
|||||
DB 367 aagaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 308
|||||
QY 181 aagaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 240
|||||
DB 307 gaaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 248
|||||
QY 241 gaaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 300
|||||
DB 247 gaaagcaatgataaagaatgagaaatccagtgatgcaacagagatcttgccaa 188
|||||
QY 301 ggcctgtagtatacattcatcagagaaagtcg 333
|||||
DB 187 ggcctgtagtatacattcatcagagaaagtcg 155
|||||

RESULT 3
AM067405 385 bp mRNA EST 13-OCT-1999
LOCUS 614041D10.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AM067405
VERSION AM067405.1 GI:6022593
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 385)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614041 row: D column: 10.
FEATURES
source location/Qualifiers
1..385
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_1lb="614 - root cDNA library from Walbot Lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="XLOLR"
/note="Organ: root; Vector: pBluescriptII SK+; Site: 1; ECORI: Site_2: XhoI; 3-4 days old root tissue from Walbot Lab (LM)"

BASE COUNT 102 a 77 c 91 g 111 t 4 others
ORIGIN

Query Match 86.7%; Score 288.8; DB 110; Length 385;
Best Local Similarity 99.3%; Pred. No. 9e-73;

Matches 290; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 42 cctcttaaggcgagatcccgctctctctgattgctcctcttgaaagcaatgaactt 101
|||||
DB 1 cctcttaaggcgagatcccgctctctctgattgctcctcttgaaagcaatgaactt 60
|||||
QY 102 gcaagcatgagctctgcaagaaggaatgataaagaatgagaaatccagtgatga 161
|||||
DB 61 gcagcctatgagccttgcgaagaaggaatgataaagaatgagaaatccagtgatga 120
|||||
QY 162 aacagagagatcttgcaagaagaagcgatgataacatgcaacttttaattttttt 221
|||||
DB 121 acacgagagatcttgcaagaagaagcgatgataacatgcaacttttaattttttt 180
|||||
QY 222 ggcctatcttgcaagaagaacagaaacatctgctctctcttgatattggaatcat 281
|||||
DB 181 ggcctatcttgcaagaagaacagaaacatctgctctctcttgatattggaatcat 240
|||||
QY 282 tcccgtagggccttgatgctgctgataacatgataacatgagaaagtcg 333
|||||
DB 241 tcccgtagggccttgatgctgctgataacatgataacatgagaaagtcg 292
|||||

RESULT 4
AM067175 552 bp mRNA EST 12-OCT-1999
LOCUS 683016G05.x1 683 - 14 day immature embryo from Hake lab (HS) Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AM067175
VERSION AM067175.1 GI:6022247
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE 1 (bases 1 to 552)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 683016 row: G column: 05.
FEATURES
source location/Qualifiers
1..552
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_1lb="683 - 14 day immature embryo from Hake lab (HS)"
/tissue_type="embryo"
/dev_stage="14 days after pollination"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pBKCMV (Stratagene's Zap Express); Site_1: XhoI; Site_2: ECORI; Directionally cloned, 14 day immature embryo library created with Stratagene's Zap Express cDNA protocol."

BASE COUNT 159 a 126 c 101 g 166 t
ORIGIN

Query Match 41.7%; Score 139; DB 110; Length 552;
Best Local Similarity 88.3%; Pred. No. 1.4e-29;
Matches 151; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 30 ggtcacctgggctcttaaggcgagatcccggtctcttgatgctcctcttgga 89
|||||

Db 545 GGATCCCTTGACTCTTTAAGGCGCGATCCCGCTTTTTCGATGCTCTTGGA 486
 Oy 90 gcaatgacttgcagctatgacttgcagagcaatgataaagaatgaaaaat 149
 Db 485 GCAATGATTTTTCAGGCTAGAGCTTCCCAAGAGCGCATCAAGAGAACAGCAGT 426
 Oy 150 tccagtagtaacaacagagatcttcccaagaagcagatgataacatg 200
 Db 425 TCCAGTAGTACACAAAGAGAACTTCCCAAGAAATCAGTATCAGAGATG 375

RESULT 5
 BE177094 499 bp mRNA EST 31-OCT-2000
 LOCUS BE177094
 DEFINITION BE1_3_F02_g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

ACCESSION BE177094
 VERSION BE177094.1 GI:11064960
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
 1 (bases 1 to 499)
 RefSeq: S.P. Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
 An EST database from Sorghum: developing embryos
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt.uga.edu

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt.uga.edu

FEATURES
 source
 1. .499
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_id="Embryo 1 (EM1)"
 /note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from lambda zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 133 a 114 c 125 g 127 t
 ORIGIN

Query Match 41.3%; Score 137.4; DB 144; Length 499;
 Best Local Similarity 87.7%; Pred. No. 4e-29;
 Matches 150; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Oy 1 tgaatacttgcacatccttgaagaggtcacttgcctttaaaggcgcgaccc 60
 Db 128 TGATAGTCTTCCATCTTGAAGATGAGGCTCCACTTGGGCTTTTAAAGGCGCGATCC 187
 Oy 61 ccgagcttcttgatgtctcctcttgagcaatgaaacttgcaggtatgacttgcca 120
 Db 188 CCCGCTTCTTGATGATGCCCTCTTGCGCAATGAACCTTGACAGGCTTGCCA 247
 Oy 121 agaagcgcaatataaagaatgagaaaattccagtgtcaacagagaga 171
 Db 248 AGAAGCGCATGATCGAAGATGAGAGCAAGTCCACAGAAATCAATAGAGAGA 298

RESULT 6
 BF656155 528 bp mRNA EST 20-DEC-2000
 LOCUS BF656155
 DEFINITION BF656155 Embryo 1 (EM1) Sorghum
 propinquum cDNA, mRNA sequence.

ACCESSION BF656155
 VERSION BF656155.1 GI:11921289
 KEYWORDS EST.
 SOURCE Sorghum propinquum.
 ORGANISM Sorghum propinquum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
 1 (bases 1 to 528)
 Cordonnier-Pratt, M.-M., Gingle, A., Sudhan, M., Marsala, C. and Pratt, L.H.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 An EST database from Sorghum: floral-induced meristems
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt.uga.edu

SEQUENCES have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: JEN REV
 High quality sequence stop: 445
 POLYA-No.

FEATURES
 source
 1. .528
 Location/Qualifiers
 /organism="Sorghum propinquum"
 /db_xref="taxon:13271"
 /clone_id="Floral-Induced Meristem 1 (FM1)"
 /note="Organ: Floral-Induced meristems; Vector: pBluescript II from lambda zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."
 BASE COUNT 140 a 118 c 136 g 134 t
 ORIGIN

Query Match 40.3%; Score 134.2; DB 151; Length 528;
 Best Local Similarity 86.5%; Pred. No. 3.4e-28;
 Matches 148; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 1 tgaatacttgcacatccttgaagaggtcacttgcctttaaaggcgcgaccc 60
 Db 112 TGATAGTCTTCCATCTTGAAGATGAGGCTCCACTTGGGCTTTTAAAGGCGCGATCC 171
 Oy 61 ccgagcttcttgatgtctcctcttgagcaatgaaacttgcaggtatgacttgcca 120
 Db 172 CCCGCTTCTTGATGATGCCCTCTTGCGCAATGAACCTTGACAGGCTTGCCA 231
 Oy 121 agaagcgcaatataaagaatgagaaaattccagtgtcaacagagaga 171
 Db 232 AGAAGCGCATGATCGAAGATGAGAGCAAGTCCACAGAAATCAATAGAGAGA 282

RESULT 7
 AV523385/c 319 bp mRNA EST 01-SEP-2000
 LOCUS AV523385
 DEFINITION AV523385 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone ApZL26a12P 3', mRNA sequence.
 ACCESSION AV523385

VERSION	AV523385.1	GI:8682913
KEYWORDS	EST.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta: eudicotyledons: core eudicots; Rosidae; eurosids II, Brassicales: Brassicaceae; Arabidops.	
AUTHORS	1 (bases 1 to 319)	
TITLE	Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.	
JOURNAL	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries	
MEDLINE	DNA Res. 7, 175-180 (2000)	
COMMENT	20363093	
FEATURES	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers	
SOURCE	1. 319 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="APZL26a12F" /clone_1bp="Arabidopsis thaliana aboveground organs two to six-week old" /library_type="aboveground organs" /day_stage="two to six-week old" /note="Vector: pluescriptII SR-; Site_1: EcoRI; Site_2: XhoI"	
BASE COUNT	99 a 86 c 55 g 79 t	
ORIGIN		
Query Match	21.5%; Score 71.6; DB 30; Length 319;	
Best Local Similarity	63.0%; Pred. No. 3.3e-10;	
Matches	98: Conservative 0; Mismatches 44; Indels 0; Gaps 0;	
QY	1 tgaatattcttccactcttgaaacgaggggtccacttgagctcttaaggcgcatcc 60	
Db	317 TGGTGTGTGATTCATTGTGCGTATGAGGACCCCTAGCTTTGTTCAAGACGACGTCC 258	
QY	61 ccgagctctcttgatgattccctctcttgagcaaatgaacttcgaagctatagacttgca 120	
Db	257 CGAGGTCTCTTCGTGGGTGCGCTCTTAGGTGCTATGAATTTGCTGGTACGAACATGACCA 198	
QY	121 agaaagcaatgataaagaatga 142	
Db	197 AGAAAGCTATGTCAGAGAAGACA 176	
RESULT	8	
LOCUS	AV523684	443 bp mRNA EST 01-SEP-2000
DEFINITION	AV523684 Arabidopsis thaliana aboveground organs two to six-week	
ACCESSION	AV523684	Old Arabidopsis thaliana cDNA clone APZL36d06F 3', mRNA sequence.
VERSION	AV523684.1	GI:8683212
KEYWORDS	EST.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
REFERENCE	Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta: eudicotyledons: core eudicots; Rosidae; eurosids II, Brassicales: Brassicaceae; Arabidops.	
AUTHORS	1 (bases 1 to 443)	
TITLE	Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.	
JOURNAL	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries	
MEDLINE	DNA Res. 7, 175-180 (2000)	
COMMENT	20363093	
	Contact: Erika Asamizu	

FEATURES					
source					
1..443					
/organism="Arabidopsis thaliana"					
/strain="Columbia"					
/db_xref="taxon:3702"					
/clone="APZL36d06F"					
/clone_1lb="Arabidopsis thaliana aboveground organs two to six-week old"					
/tissue_type="aboveground organs"					
/dev_stage="two to six-week old"					
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"					
BASE COUNT					
127 a 125 c 77 g 114 t					
ORIGIN					
Query Match 21.5%; Score 71.6; DB 30; Length 443;					
Best Local Similarity 69.0%; Pred. No. 3.6e-10;					
Matches 98; Conservative 0; Mismatches 44; Indels 0; Gaps 0;					
OY	1	tgaatactgtccatccttgaaacgagggccactggcgctcttaaggcgcatcc	60		
Db	330	TGCTGTGTATCGATTCCTCCGTAATGAGGCACGCCCTTGTTTCAAGACGACATCC	271		
OY	61	cgcgcttccttgatgatgtcctctcttgagaagaataacttcgagcgcatgagctgcga	120		
Db	270	CGAGGTTCTTGGGTGCGCTCCTCTCAGGTGCTATGAACATTGTGCGTAGCAACTAGCCA	211		
OY	121	agaagcgaatgataaagatatga	142		
Db	210	AGAAAGCATATCAGACAGAACGA	189		
RESULT 9					
AV561206/c mRNA EST 07-SEP-2000					
LOCUS AV561206 Arabidopsis thaliana green siliques Columbia Arabidopsis					
DEFINITION thaliana cDNA clone SQ147e05f 3', mRNA sequence.					
ACCESSION AV561206					
VERSION AV561206					
KEYWORDS AV561206.1 GI:8732632					
SOURCE EST.					
ORGANISM thale cress.					
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;					
Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;					
Brassicales; Brassicaceae; Arabidopsiis.					
1 (bases 1 to 572)					
Asamizu E., Nakamura Y., Sato S. and Tabata S.					
A large scale analysis of cDNA in Arabidopsis thaliana: Generation					
of 12,028 non-redundant expressed sequence tags from normalized and					
size-selected cDNA libraries					
DNA Res. 7, 175-180 (2000)					
CONTACT: Erika Asamizu					
The First Laboratory for Plant Gene Research					
Kazusa DNA Research Institute					
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan					
Email: asamizue@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.					
Location/Qualifiers					
1..572					
/organism="Arabidopsis thaliana"					
/strain="Columbia"					
/db_xref="taxon:3702"					
/clone="SQ147e05f"					
/clone_1lb="Arabidopsis thaliana green siliques Columbia"					
/tissue_type="green siliques"					
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:					
XhoI"					
BASE COUNT					
158 a 161 c 102 g 151 t					
FEATURES					
SOURCE					
1..572					
/organism="Arabidopsis thaliana"					
/strain="Columbia"					
/db_xref="taxon:3702"					
/clone="SQ147e05f"					
/clone_1lb="Arabidopsis thaliana green siliques Columbia"					
/tissue_type="green siliques"					
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:					
XhoI"					
BASE COUNT					
158 a 161 c 102 g 151 t					

VERSION BF278348.1 GI:11209418
 KEYWORDS EST.
 SOURCE Gossypium arboreum.
 ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 468)
 AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber
 JOURNAL Unpublished (2000)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4233
 Email: rwing@clemson.edu
 Seq primer: TAATACGACTCACTATAGG
 High quality sequence start: 6
 High quality sequence stop: 462.
 FEATURES
 Location/Qualifiers
 1..468
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_EB0032H10f"
 /clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
 /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
 /lab_host="E. coli"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 140 a 84 c 92 g 152 t
 ORIGIN
 Query Match 20.3%; Score 67.6; DB 146; Length 468;
 Best Local Similarity 64.9%; Pred. No. 5.1e-09;
 Matches 100; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
 Oy 4 taatctgcatcctctggaacgaggtccacttggtctcttaaggcgcatcccg 63
 Db 9 TAGCCTTTCTACTCGTCATGAGGACCCCTTGCGCTTCAAGAGCGCTACCGA 68
 Oy 64 gctctcttgatctctctctctgagcaatgaacttgcaagctatgagcttgcaca 123
 Db 69 GCTTTCTGGATTGCTCCCTTAGGTGCATGAACCTTCCGCGTACGAGTTGTGAGCA 128
 Oy 124 aggcacatataaagatgagaaaattccagctga 157
 Db 129 AGGCCATGAGCGAACAAGATGTTGCTAGGCA 162
 RESULT 13
 BE642959 790 bp mRNA EST 01-SEP-2000
 LOCUS Cr12.7_H22-SP6 Ceratopteris Spore Library Ceratopteris richardii
 DEFINITION cDNA clone Cr12.7_H22 5', mRNA sequence.
 ACCESSION BE642959
 VERSION BE642959.1 GI:9960640
 KEYWORDS EST.
 SOURCE Ceratopteris richardii.
 ORGANISM Ceratopteris richardii.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Filicophyta; Filicopsida; Filicales; Pteridaceae; Ceratopteris.
 REFERENCE 1 (bases 1 to 790)
 AUTHORS Chatterjee, A., San Miguel, P., Stout, S.C., Banks, J. and Roux, S.J.
 TITLE Expressed sequence tags of cDNA clones from a C. richardii library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Roux SJ

Section of Molecular Cell and Developmental Biology
 University of Texas
 Biology Building, Room 16, Austin, TX 78712, USA
 Tel: 512 471 4238
 Fax: 512 232 3402
 Email: sroux@uts.cc.utexas.edu
 Plate: Cr12.7 row: H column: 22
 Seq primer: SP6.
 FEATURES
 Location/Qualifiers
 1..790
 /organism="Ceratopteris richardii"
 /cultivar="Brogan"
 /db_xref="taxon:49495"
 /clone="Cr12.7_H22"
 /clone_1lb="Ceratopteris Spore Library"
 /tissue_type="Gametophyte"
 /cell_type="Spore"
 /dev_stage="20 hours after germination initiation"
 /note="Vector: pCMVSPORT6; EST sequence from cDNA library.
 cDNA library constructed from mRNA isolated from C. richardii spores that had developed for 20 hours after their germination had been initiated by white light."
 BASE COUNT 215 a 137 c 206 g 232 t
 ORIGIN
 Query Match 19.7%; Score 65.6; DB 137; Length 790;
 Best Local Similarity 63.1%; Pred. No. 2.2e-08;
 Matches 101; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 Oy 9 ttgtccatccttggaacgaggtccacttggctcttaaggcgcatcccgcttc 68
 Db 171 TGCACCATGTGAAGAGAGAGCGCTTGCTAGGTATTCAAAGGTGCAAGTTC 230
 Oy 69 tcttgatgctccctcttgagcaatgaacttgagctatgagcttgcacaagagca 128
 Db 231 TTCTGATTGCACCTCTGAGCAATGAATTTCTCTGGGTGCAACTGCAAAAGAGCC 290
 Oy 129 atgataaagatgagaaaattccagtgaatcaacagag 168
 Db 291 ATGGAATAGAAAGACGAGATTGACTGGCTGTCAAGTG 330
 RESULT 14
 BE919380 367 bp mRNA EST 02-OCT-2000
 LOCUS EST423233 potato leaves and petioles Solanum tuberosum cDNA clone
 DEFINITION EST423233 potato leaves and petioles Solanum tuberosum cDNA clone
 ACCESSION BE919380
 VERSION BE919380.1 GI:10445540
 KEYWORDS EST.
 SOURCE potato.
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 REFERENCE 1 (bases 1 to 367)
 AUTHORS van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J., Uteback, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R., Ronning, C.M., Fry, W.E., Tanksley, S.D. and Baker, B.
 TITLE Generation of ESTs from potato leaves and petioles
 JOURNAL The Institute for Genomic Research
 COMMENT For clone request: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com.
 FEATURES
 Location/Qualifiers
 1..367
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="cSTB1H19"
 /clone_1lb="potato leaves and petioles"
